

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> A61K 39/12, C07H 21/04 C07K 7/06, 7/08, 7/10 C07K 13/00, 15/28	<b>A1</b>	<b>(11) International Publication Number:</b> WO 90/11089 <b>(43) International Publication Date:</b> 4 October 1990 (04.10.90)									
<b>(21) International Application Number:</b> PCT/US90/01348 <b>(22) International Filing Date:</b> 15 March 1990 (15.03.90)  <b>(30) Priority data:</b> <table border="0" style="width: 100%;"><tr><td style="width: 30%;">325,338</td><td style="width: 40%;">17 March 1989 (17.03.89)</td><td style="width: 30%;">US</td></tr><tr><td>341,334</td><td>20 April 1989 (20.04.89)</td><td>US</td></tr><tr><td>355,002</td><td>18 May 1989 (18.05.89)</td><td>US</td></tr></table> <b>(71) Applicant:</b> CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).  <b>(72) Inventors:</b> HOUGHTON, Michael ; 53 Rosemead Court, Danville, CA 94526 (US). CHOO, Qui-Lim ; 5700 Fern Street, El Cerrito, CA 94530 (US). KUO, George ; 1370 Sixth Avenue, San Francisco, CA 94122 (US).		325,338	17 March 1989 (17.03.89)	US	341,334	20 April 1989 (20.04.89)	US	355,002	18 May 1989 (18.05.89)	US	<b>(74) Agents:</b> MONROY, Gladys, H. et al.; Irell & Manella, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).  <b>(81) Designated States:</b> AU, FI, HU, JP, KR, NO, SU.  <b>Published</b> <i>With international search report.</i>
325,338	17 March 1989 (17.03.89)	US									
341,334	20 April 1989 (20.04.89)	US									
355,002	18 May 1989 (18.05.89)	US									
<b>(54) Title:</b> NANBV DIAGNOSTICS AND VACCINES  <b>(57) Abstract</b> <p>A new virus, Hepatitis C virus (HCV), which has proven to be the major etiologic agent of blood-borne NANBH, was discovered by Applicant. The initial work on this virus, which includes a partial genomic sequence of the prototype HCV isolate, is described in EPO Pub. No. 318,216, and PCT Pub. No. WO 89/04669. The present invention, which in part is based on new HCV sequences and polypeptides which are not disclosed in the above-cited publications, includes the application of these new sequences and polypeptides in immunoassays, probe diagnostics, anti-HCV antibody production, PCR technology, and recombinant DNA technology. Included within the invention also are novel, immunogenic polypeptides encoded within clones containing HCV cDNA, novel methods for purifying an immunogenic HCV polypeptide, and antisense polynucleotides derived from HCV cDNA.</p>											

#### DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

#### *FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic	KR	Republic of Korea	SN	Senegal
CG	Congo	LJ	Liechtenstein	SU	Soviet Union
CH	Switzerland	LK	Sri Lanka	TD	Chad
CM	Cameroon	LU	Luxembourg	TG	Togo
DE	Germany, Federal Republic of	MC	Monaco	US	United States of America
DK	Denmark				

5

NANBV DIAGNOSTICS AND VACCINES10 Technical Field

The invention relates to materials and methodologies for managing the spread of non-A, non-B hepatitis virus (NANBV) infection. More specifically, it relates to polynucleotides derived from the genome of an  
15 etiologic agent of NANBH, hepatitis C virus (HCV), to polypeptides encoded therein, and to antibodies directed to the polypeptides. These reagents are useful as screening agents for HCV and its infection, and as protective  
20 agents against the disease.

References Cited in the Application

- Barr et al. (1986), Biotechniques 4:428.  
Botstein (1979), Gene 8:17.  
25 Brinton, M.A. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.327-374.  
Broach (1981) in: Molecular Biology of the Yeast  
30 Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.  
Broach et al. (1983), Meth. Enz. 101:307.  
Chang et al. (1977), Nature 198:1056.  
Chirgwin et al. (1979), Biochemistry 18:5294.  
Chomczynski and Sacchi (1987), Analytical Biochemistry  
35 162:156.  
Clewell et al. (1969), Proc. Natl. Acad. Sci. USA 62:1159.

- Clewell (1972), J. Bacteriol. 110:667.
- Cohen (1972), Proc. Natl. Acad. Sci. USA 69:2110.
- Cousens et al. (1987), gene 61:265.
- 5 De Boer et al. (1983), Proc. Natl. Acad. Sci. USA 292:128.
- Dreesman et al. (1985), J. Infect. Disease 151:761.
- Feinstone, S.M. and Hoofnagle, J.H. (1984), New Engl. J. Med. 311:185.
- 10 Fields & Knipe (1986), FUNDAMENTAL VIROLOGY (Raven Press, N.Y.).
- Fiers et al. (1978), Nature 273:113.
- Gerety, R.J. et al., in VIRAL HEPATITIS AND LIVER DISEASE (Vyas, B.N., Dienstag, J.L., and Hoofnagle, J.H., eds, Grune and Stratton, Inc., 1984) pp 23-47.
- 15 Goeddel et al. (1980), Nucleic Acids Res. 8:4057.
- Graham and Van der Eb (1978), Virology 52:546.
- Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961.
- Grych et al. (1985), Nature 316:74.
- 20 Gubler and Hoffman (1983), Gene 25:263.
- Hahn (1988) Virology 162:167.
- Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS.
- Han (1987), Biochemistry 26:1617.
- 25 Helfman (1983), Proc. Natl. Acad. Sci. USA 80:31.
- Hess et al. (1968), J. Adv. Enzyme Reg 7:149.
- Hinnen et al. (1978), Proc. Natl. Acad. Sci. 75:1929.
- Hitzeman et al. (1980), J. Biol. Chem. 255:2073.
- Holland et al. (1978), Biochemistry 17:4900.
- 30 Holland (1981), J. Biol. Chem. 256: 1385.
- Houghton et al. (1981), Nucleic Acids Res. 9:247
- Hunyh, T.V. et al. (1985) in DNA CLONING TECHNIQUES; A PRACTICAL APPROACH (D. Glover, Ed., IRL Press, Oxford, U.K.) pp. 49-78.
- 35 Immun. Rev. (1982) 62:185.
- Iwarson (1987), British Medical J. 295:946.

- Kennett et al. (1980) MONOCLONAL ANTIBODIES.  
Kyte and Doolittle (1982), J. Mol. Biol. 157:105-132.  
Laemmli (1970), Nature 227, 680.  
5 Lee et al. (1988), Science 239:1288.  
Maniatis, T., et al. (1982) MOLECULAR CLONING; A  
LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring  
Harbor, N.Y.).  
10 Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN  
CELL AND MOLECULAR BIOLOGY (Academic Press, London).  
Maxam et al. (1980), Methods in Enzymology 65:499.  
MacNamara et al. (1984), Science 226:1325.  
Messing et al. (1981), Nucleic Acids Res. 9:309.  
Messing (1983), Methods in Enzymology 101:20-37.  
15 METHODS IN ENZYMOLOGY (Academic Press).  
Michelle et al., Int. Symposium on Viral Hepatitis.  
Monath (1986) in THE VIRUSES: THE TOGAVIRIDAE AND  
FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol.  
eds. Schlesinger and Schlesinger, Plenum Press), p.375-  
20 440.  
Nagahuma et al. (1984), Anal. Biochem. 141:74.  
Neurath et al. (1984), Science 224:392.  
Nisonoff et al. (1981), Clin. Immunol. Immunopathol.  
21:397-406.  
25 Overby, L.R. (1985), Curr. Hepatol. 5:49.  
Overby, L.R. (1986), Curr. Hepatol. 6:65.  
Overby, L.R. (1987), Curr. Hepatol. 7:35.  
Peleg (1969), Nature 221:193.  
Pfefferkorn and Shapiro (1974), in COMPREHENSIVE VIROLOGY,  
30 Vol. 2 (Fraenkel-Conrat & Wagner, eds., Plenum, N.Y.) pp.  
171-230.  
Prince, A.M. (1983), Annu. Rev. Microbiol. 37:217.  
Rice et al. (1985), Science 229:726.

- Rice et al. (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press), p.279-328.
- Roehrig (1986) in THE VIRUSES: THE TOGAVIRIDAE AND FLAVIVIRIDAE (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press)
- Sadler et al. (1980), Gene 8, 279.
- Saiki et al. (1986), Nature 324: 163.
- Saiki et al. (1988), Science 239:487.
- Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463.
- Schlesinger et al. (1986), J. Virol. 60:1153.
- Schreier, M., et al. (1980) HYBRIDOMA TECHNIQUES
- Scopes (1984), PROTEIN PURIFICATION, PRINCIPLES AND PRACTICE, SECOND EDITION (Springer-Verlag, N.Y.).
- Shimatake et al. (1981), Nature 292:128.
- Sippel (1973), Eur. J. Biochem. 37:31.
- Steimer et al. (1986), J. Virol. 58:9.
- Stollar (1980), in THE TOGAVIRUSES (R.W. Schlesinger, ed., Academic Press, N.Y.), pp. 584-622.
- Sumiyoshi et al. (1987), Virology 161:497.
- Taylor et al. (1976), Biochem. Biophys. Acta 442:324.
- Towbin et al. (1979), Proc. Natl. Acad. Sci. USA 76, 4350.
- Tsu and Herzenberg (1980), in SELECTED METHODS IN CELLULAR IMMUNOLOGY (W.H. Freeman and Co.) pp. 373-391.
- Vytdehaag et al. (1985), J. Immunol. 134:1225.
- Valenzuela, P., et al. (1982), Nature 298:344.
- Valenzuela, P., et al. (1984), in HEPATITIS B (Millman, I., et al., ed, Plenum Press) pp. 225-236.
- Warner (1984), DNA 3:401.
- Wu and Grossman (1987), Methods in Enzymology Vol. 154, RECOMBINANT DNA, Part E.
- Wu (1987), Methods in Enzymology vol 155, RECOMBINANT DNA, part F.
- Zoller (1982), Nucleic Acids Res. 10:6487.

Cited Patents

- 5 EPO Pub. No. 318,216  
PCT Pub. No. WO 89/04669  
U.S. Patent No. 4,341,761  
U.S. Patent No. 4,399,121  
U.S. Patent No. 4,427,783  
U.S. Patent No. 4,444,887  
10 U.S. Patent No. 4,466,917  
U.S. Patent No. 4,472,500  
U.S. Patent No. 4,491,632  
U.S. Patent No. 4,493,890

Background Art

- 15 Non-A, Non-B hepatitis (NANBH) is a transmissible disease or family of diseases that are believed to be viral-induced, and that are distinguishable from other forms of viral-associated liver diseases, including that caused by the known hepatitis viruses,  
20 i.e., hepatitis A virus (HAV), hepatitis B virus (HBV), and delta hepatitis virus (HDV), as well as the hepatitis induced by cytomegalovirus (CMV) or Epstein-Barr virus (EBV). NANBH was first identified in transfused individuals. Transmission from man to chimpanzee and se-  
25 rial passage in chimpanzees provided evidence that NANBH is due to a transmissible infectious agent or agents.

- Epidemiologic evidence is suggestive that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the  
30 sporadically occurring (community acquired) type. However, the number of agents which may be the causative of NANBH are unknown.

- Clinical diagnosis and identification of NANBH has been accomplished primarily by exclusion of other  
35 viral markers. Among the methods used to detect putative NANBV antigens and antibodies are agar-gel diffusion,

counterimmunoelectrophoresis, immunofluorescence  
microscopy, immune electron microscopy, radioimmunoassay,  
and enzyme-linked immunosorbent assay. However, none of  
5 these assays has proved to be sufficiently sensitive,  
specific, and reproducible to be used as a diagnostic test  
for NANBH.

Previously there was neither clarity nor agree-  
ment as to the identity or specificity of the antigen  
10 antibody systems associated with agents of NANBH. This  
was due, at least in part, to the prior or co-infection of  
HBV with NANBV in individuals, and to the known complexity  
of the soluble and particulate antigens associated with  
HBV, as well as to the integration of HBV DNA into the  
15 genome of liver cells. In addition, there is the pos-  
sibility that NANBH is caused by more than one infectious  
agent, as well as the possibility that NANBH has been mis-  
diagnosed. Moreover, it is unclear what the serological  
assays detect in the serum of patients with NANBH. It has  
20 been postulated that the agar-gel diffusion and counter-  
immunoelectrophoresis assays detect autoimmune responses  
or nonspecific protein interactions that sometimes occur  
between serum specimens, and that they do not represent  
specific NANBV antigen-antibody reactions. The  
25 immunofluorescence, and enzyme-linked immunosorbent, and  
radioimmunoassays appear to detect low levels of a  
rheumatoid-factor-like material that is frequently present  
in the serum of patients with NANBH as well as in patients  
with other hepatic and nonhepatic diseases. Some of the  
30 reactivity detected may represent antibody to host-  
determined cytoplasmic antigens.

There have been a number of candidate NANBV.  
See, for example the reviews by Prince (1983), Feinstone  
and Hoofnagle (1984), and Overby (1985, 1986, 1987) and  
35 the article by Iwarson (1987). However, there is no proof



that any of these candidates represent the etiological agent of NANBH.

5       The demand for sensitive, specific methods for screening and identifying carriers of NANBV and NANBV contaminated blood or blood products is significant. Post-transfusion hepatitis (PTH) occurs in approximately 10% of transfused patients, and NANBH accounts for up to 90% of these cases. The major problem in this disease is 10 the frequent progression to chronic liver damage (25-55%).

      Patient care as well as the prevention of transmission of NANBH by blood and blood products or by close personal contact require reliable screening, 15 diagnostic and prognostic tools to detect nucleic acids, antigens and antibodies related to NANBV. In addition, there is also a need for effective vaccines and immunotherapeutic therapeutic agents for the prevention and/or treatment of the disease.

20       Applicant discovered a new virus, the Hepatitis C virus (HCV), which has proven to be the major etiologic agent of blood-borne NANBH (BB-NANBH). Applicant's initial work, including a partial genomic sequence of the prototype HCV isolate, CDC/HCV1 (also called HCV1), is described in EPO Pub. No. 318,216 (published 31 May 1989) 25 and PCT Pub. No. WO 89/04669 (published 1 June 1989). The disclosures of these patent applications, as well as any corresponding national patent applications, are incorporated herein by reference. These applications teach, inter alia, recombinant DNA methods of cloning and 30 expressing HCV sequences, HCV polypeptides, HCV immunodiagnostic techniques, HCV probe diagnostic techniques, anti-HCV antibodies, and methods of isolating new hCV sequences, including sequences of new HCV isolates.

35

Disclosure of the Invention

5 The present invention is based, in part, on new HCV sequences and polypeptides that are not disclosed in EPO Pub. No. 318,216, or in PCT Pub. No. WO 89/04669. Included within the invention is the application of these new sequences and polypeptides in, inter alia, immunodiagnos-  
10 tics, probe diagnostics, anti-HCV antibody production, PCR technology and recombinant DNA technology. Included within the invention, also, are new immunoassays based upon the immunogenicity of HCV polypeptides disclosed herein. The new subject matter claimed herein, while developed using techniques described in, for  
15 example, EPO Pub. No. 318,216, has a priority date which antecedes that publication, or any counterpart thereof. Thus, the invention provides novel compositions and methods useful for screening samples for HCV antigens and antibodies, and useful for treatment of HCV infections.

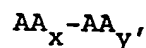
20 Accordingly, one aspect of the invention is a recombinant polynucleotide comprising a sequence derived from HCV cDNA, wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone  
25 CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

30 Another aspect of the invention is a purified polypeptide comprising an epitope encoded within HCV cDNA wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

35 Yet another aspect of the invention is an immunogenic polypeptide produced by a cell transformed with a recombinant expression vector comprising an ORF of DNA derived from HCV cDNA, wherein the HCV cDNA is comprised

of a sequence derived from the HCV cDNA sequence in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or  
 5 clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the ORF is operably linked to a control sequence compatible with a desired host.

Another aspect of the invention is a peptide  
 10 comprising an HCV epitope, wherein the peptide is of the formula



15 wherein x and y designate amino acid numbers shown in Fig. 17, and wherein the peptide is selected from the group consisting of AA1-AA25, AA1-AA50, AA1-AA84, AA9-AA177, AA1-AA10, AA5-AA20, AA20-AA25, AA35-AA45, AA50-AA100, AA40-AA90, AA45-AA65, AA65-AA75, AA80-90, AA99-AA120,  
 20 AA95-AA110, AA105-AA120, AA100-AA150, AA150-AA200, AA155-AA170, AA190-AA210, AA200-AA250, AA220-AA240, AA245-AA265, AA250-AA300, AA290-AA330, AA290-305, AA300-AA350, AA310-AA330, AA350-AA400, AA380-AA395, AA405-AA495, AA400-AA450, AA405-AA415, AA415-AA425,  
 25 AA425-AA435, AA437-AA582, AA450-AA500, AA440-AA460, AA460-AA470, AA475-AA495, AA500-AA550, AA511-AA690, AA515-AA550, AA550-AA600, AA550-AA625, AA575-AA605, AA585-AA600, AA600-AA650, AA600-AA625, AA635-AA665, AA650-AA700, AA645-AA680, AA700-AA750, AA700-AA725, AA700-AA750,  
 30 AA725-AA775, AA770-AA790, AA750-AA800, AA800-AA815, AA825-AA850, AA850-AA875, AA800-AA850, AA920-AA990, AA850-AA900, AA920-AA945, AA940-AA965, AA970-AA990, AA950-AA1000, AA1000-AA1060, AA1000-AA1025, AA1000-AA1050, AA1025-AA1040, AA1040-AA1055, AA1075-AA1175,  
 35 AA1050-AA1200, AA1070-AA1100, AA1100-AA1130, AA1140-AA1165, AA1192-AA1457, AA1195-AA1250,

AA1200-AA1225, AA1225-AA1250, AA1250-AA1300,  
AA1260-AA1310, AA1260-AA1280, AA1266-AA1428,  
AA1300-AA1350, AA1290-AA1310, AA1310-AA1340, AA1345-  
5 AA1405, AA1345-AA1365, AA1350-AA1400, AA1365-AA1380,  
AA1380-AA1405, AA1400-AA1450, AA1450-AA1500,  
AA1460-AA1475, AA1475-AA1515, AA1475-AA1500,  
AA1500-AA1550, AA1500-AA1515, AA1515-AA1550,  
AA1550-AA1600, AA1545-AA1560, AA1569-AA1931,  
10 AA1570-AA1590, AA1595-AA1610, AA1590-AA1650,  
AA1610-AA1645, AA1650-AA1690, AA1685-AA1770,  
AA1689-AA1805, AA1690-AA1720, AA1694-AA1735,  
AA1720-AA1745, AA1745-AA1770, AA1750-AA1800,  
AA1775-AA1810, AA1795-AA1850, AA1850-AA1900,  
15 AA1900-AA1950, AA1900-AA1920, AA1916-AA2021,  
AA1920-AA1940, AA1949-AA2124, AA1950-AA2000,  
AA1950-AA1985, AA1980-AA2000, AA2000-AA2050,  
AA2005-AA2025, AA2020-AA2045, AA2045-AA2100,  
AA2045-AA2070, AA2054-AA2223, AA2070-AA2100,  
20 AA2100-AA2150, AA2150-AA2200, AA2200-AA2250,  
AA2200-AA2325, AA2250-AA2330, AA2255-AA2270,  
AA2265-AA2280, AA2280-AA2290, AA2287-AA2385,  
AA2300-AA2350, AA2290-AA2310, AA2310-AA2330,  
AA2330-AA2350, AA2350-AA2400, AA2348-AA2464,  
25 AA2345-AA2415, AA2345-AA2375, AA2370-AA2410,  
AA2371-AA2502, AA2400-AA2450, AA2400-AA2425,  
AA2415-AA2450, AA2445-AA2500, AA2445-AA2475,  
AA2470-AA2490, AA2500-AA2550, AA2505-AA2540,  
AA2535-AA2560, AA2550-AA2600, AA2560-AA2580,  
30 AA2600-AA2650, AA2605-AA2620, AA2620-AA2650,  
AA2640-AA2660, AA2650-AA2700, AA2655-AA2670,  
AA2670-AA2700, AA2700-AA2750, AA2740-AA2760,  
AA2750-AA2800, AA2755-AA2780,  
AA2780-AA2830, AA2785-AA2810, AA2796-AA2886,  
35 AA2810-AA2825, AA2800-AA2850, AA2850-AA2900,

AA2850-AA2865, AA2885-AA2905, AA2900-AA2950,  
AA2910-AA2930, AA2925-AA2950, AA2945-end(C' terminal).

5 Still another aspect of the invention is a  
monoclonal antibody directed against an epitope encoded in  
HCV cDNA, wherein the HCV cDNA is of a sequence indicated  
by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig.  
17, or is the sequence present in clone 13i, or clone 26j,  
or clone 59a, or clone 84a, or clone CA156e, or clone  
10 167b, or clone pil4a, or clone CA216a, or clone CA290a, or  
clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Yet another aspect of the invention is a  
preparation of purified polyclonal antibodies directed  
against a polypeptide comprised of an epitope encoded  
15 within HCV cDNA, wherein the HCV cDNA is of a sequence  
indicated by nucleotide numbers -319 to 1348 or 8659 to  
8866 in Fig. 17, or is the sequence present in in clone  
13i, or clone 26j, or clone 59a, or clone 84a, or clone  
CA156e, or clone 167b, or clone pil4a, or clone CA216a, or  
20 clone CA290a, or clone ag30a, or clone 205a, or clone 18g,  
or clone 16jh.

Still another aspect of the invention is a  
polynucleotide probe for HCV, wherein the probe is  
comprised of an HCV sequence derived from an HCV cDNA  
25 sequence indicated by nucleotide numbers -319 to 1348 or  
8659 to 8866 in Fig. 17, or from the complement of the HCV  
cDNA sequence.

Yet another aspect of the invention is a kit for  
analyzing samples for the presence of polynucleotides from  
30 HCV comprising a polynucleotide probe containing a  
nucleotide sequence of about 8 or more nucleotides,  
wherein the nucleotide sequence is derived from HCV cDNA  
which is of a sequence indicated by nucleotide numbers -  
319 to 1348 or 8659 to 8866 in Fig. 17, wherein the  
35 polynucleotide probe is in a suitable container.

Another aspect of the invention is a kit for analyzing samples for the presence of an HCV antigen comprising an antibody which reacts immunologically with an HCV antigen, wherein the antigen contains an epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone p114a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Yet another aspect of the invention is a kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide containing an HCV epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone p114a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

Another aspect of the invention is a kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide expressed from HCV cDNA in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, wherein the antigenic polypeptide is present in a suitable container.

Still another aspect of the invention is a method for detecting HCV nucleic acids in a sample comprising:

(a) reacting nucleic acids of the sample with a polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, and wherein the

reacting is under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample; and (b) detecting a  
5 polynucleotide duplex which contains the probe, formed in step (a).

Yet another aspect of the invention is an immunoassay for detecting an HCV antigen comprising:

(a) incubating a sample suspected of containing  
10 an HCV antigen with an antibody directed against an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or  
15 clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and (b) detecting an antibody-antigen  
20 complex formed in step (a) which contains the antibody.

Still another aspect of the invention is an immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing  
25 anti-HCV antibodies with an antigen polypeptide containing an epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or  
30 clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and detecting an antibody-antigen  
35 complex formed in step (a) which contains the antigen polypeptide.

Another aspect of the invention is a vaccine for treatment of HCV infection comprising an immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17 or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone p114a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

Yet another aspect of the invention is a method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is of the sequence present in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

Still another aspect of the invention is an antisense polynucleotide derived from HCV cDNA, wherein the HCV cDNA is that shown in Fig. 17.

Yet another aspect of the invention is a method for preparing purified fusion polypeptide C100-3 comprising:

(a) providing a crude cell lysate containing polypeptide C100-3,



- (b) treating the crude cell lysate with an amount of acetone which causes the polypeptide to precipitate,
- 5 (c) isolating and solubilizing the precipitated material,
- (d) isolating the C100-3 polypeptide by anion exchange chromatography, and
- 10 (e) further isolating the C100-3 polypeptide of step (d) by gel filtration.

#### Brief Description of the Drawings

15 Fig. 1 shows the sequence of the HCV cDNA in clone 12f, and the amino acids encoded therein.

Fig. 2 shows the HCV cDNA sequence in clone k9-1, and the amino acids encoded therein.

20 Fig. 3 shows the sequence of clone 15e, and the amino acids encoded therein.

Fig. 4 shows the nucleotide sequence of HCV cDNA in clone 13i, the amino acids encoded therein, and the sequences which overlap with clone 12f.

25 Fig. 5 shows the nucleotide sequence of HCV cDNA in clone 26j, the amino acids encoded therein, and the sequences which overlap clone 13i.

Fig. 6 shows the nucleotide sequence of HCV cDNA in clone CA59a, the amino acids encoded therein, and the sequences which overlap with clones 26j and K9-1.

30 Fig. 7 shows the nucleotide sequence of HCV cDNA in clone CA84a, the amino acids encoded therein, and the sequences which overlap with clone CA59a.

35 Fig. 8 shows the nucleotide sequence of HCV cDNA in clone CA156e, the amino acids encoded therein, and the sequences which overlap with CA84a.

Fig. 9 shows the nucleotide sequence of HCV cDNA in clone CA167b, the amino acids encoded therein, and the sequences which overlap CA156e.

Fig. 10 shows the nucleotide sequence of HCV cDNA in clone CA216a, the amino acids encoded therein, and the overlap with clone CA167b.

Fig. 11 shows the nucleotide sequence of HCV cDNA in clone CA290a, the amino acids encoded therein, and the overlap with clone CA216a.

Fig. 12 shows the nucleotide sequence of HCV cDNA in clone ag30a and the overlap with clone CA290a.

Fig. 13 shows the nucleotide sequence of HCV cDNA in clone CA205a, and the overlap with the HCV cDNA sequence in clone CA290a.

Fig. 14 shows the nucleotide sequence of HCV cDNA in clone 18g, and the overlap with the HCV cDNA sequence in clone ag30a.

Fig. 15 shows the nucleotide sequence of HCV cDNA in clone 16jh, the amino acids encoded therein, and the overlap of nucleotides with the HCV cDNA sequence in clone 15e.

Fig. 16 shows the ORF of HCV cDNA derived from clones pil4a, CA167b, CA156e, CA84a, CA59a, K9-1, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, and 15e.

Fig. 17 shows the sense strand of the compiled HCV cDNA sequence derived from the above-described clones and the compiled HCV cDNA sequence published in EPO Pub. No. 318,216. The clones from which the sequence was derived are pil4a, 18g, ag30a, CA205a, CA290a, CA216a, pil4a, CA167b, CA156e, CA84a, CA59a, K9-1 (also called k9-1), 26j, 13i, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a, and 16jh. In the figure the three horizontal dashes above the sequence indicate the position

of the putative initiator methionine codon; the two vertical dashes indicate the first and last nucleotides of the published sequence. Also shown in the figure is the amino acid sequence of the putative polyprotein encoded in the HCV cDNA.

Fig. 18 is a diagram of the immunological colony screening method used in antigenic mapping studies.

Fig. 19 shows the hydrophobicity profiles of polyproteins encoded in HCV and in West Nile virus.

Fig. 20 is a tracing of the hydrophilicity/hydrophobicity profile and of the antigenic index of the putative HCV polyprotein.

Fig. 21 shows the conserved co-linear peptides in HCV and Flaviviruses.

#### Modes for Carrying Out the Invention

##### I. Definitions

The term "hepatitis C virus" has been reserved by workers in the field for an heretofore unknown etiologic agent of NANBH. Accordingly, as used herein, "hepatitis C virus" (HCV) refers to an agent causative of NANBH, which was formerly referred to as NANBV and/or BB-NANBV. The terms HCV, NANBV, and BB-NANBV are used interchangeably herein. As an extension of this terminology, the disease caused by HCV, formerly called NANB hepatitis (NANBH), is called hepatitis C. The terms NANBH and hepatitis C may be used interchangeably herein.

The term "HCV", as used herein, denotes a viral species of which pathogenic strains cause NANBH, and attenuated strains or defective interfering particles derived therefrom. As shown infra., the HCV genome is comprised of RNA. It is known that RNA containing viruses have relatively high rates of spontaneous mutation, i.e., reportedly on the order of  $10^{-3}$  to  $10^{-4}$  per incorporated

nucleotide (Fields & Knipe (1986)). Therefore, there are multiple strains, which may be virulent or avirulent, within the HCV species described infra. The compositions and methods described herein, enable the propagation, identification, detection, and isolation of the various HCV strains or isolates. Moreover, the disclosure herein allows the preparation of diagnostics and vaccines for the various strains, as well as compositions and methods that have utility in screening procedures for anti-viral agents for pharmacologic use, such as agents that inhibit replication of HCV.

The information provided herein, although derived from the prototype strain or isolate of HCV, hereinafter referred to as CDC/HCV1 (also called HCV1), is sufficient to allow a viral taxonomist to identify other strains which fall within the species. The information provided herein allows the belief that HCV is a Flavi-like virus. The morphology and composition of Flavivirus particles are known, and are discussed in Brinton (1986). Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections that are about 5-10 nm long with terminal knobs about 2 nm in diameter.

Different strains or isolates of HCV are expected to contain variations at the amino acid and nucleic acids compared with the prototype isolate, HCV1. Many isolates are expected to show much (i.e. more than about 40%) homology in the total amino acid sequence compared with HCV1. However, it may also be found that other less homologous HCV isolates. These would be defined as HCV strains according to various criteria such as an ORF of approximately 9,000 nucleotides to ap-

proximately 12,000 nucleotides, encoding a polyprotein similar in size to that of HCV1, an encoded polyprotein of similar hydrophobic and antigenic character to that of HCV1, and the presence of co-linear peptide sequences that are conserved with HCV1. In addition, the genome would be a positive-stranded RNA.

HCV encodes at least one epitope which is immunologically identifiable with an epitope in the HCV genome from which the cDNAs described herein are derived; preferably the epitope is contained an amino acid sequence described herein. The epitope is unique to HCV when compared to other known Flaviviruses. The uniqueness of the epitope may be determined by its immunological reactivity with anti-HCV antibodies and lack of immunological reactivity with antibodies to other Flavivirus species. Methods for determining immunological reactivity are known in the art, for example, by radioimmunoassay, by Elisa assay, by hemagglutination, and several examples of suitable techniques for assays are provided herein.

In addition to the above, the following parameters of nucleic acid homology and amino acid homology are applicable, either alone or in combination, in identifying a strain or isolate as HCV. Since HCV strains and isolates are evolutionarily related, it is expected that the overall homology of the genomes at the nucleotide level probably will be about 40% or greater, probably about 60% or greater, and even more probably about 80% or greater; and in addition that there will be corresponding contiguous sequences of at least about 13 nucleotides. The correspondence between the putative HCV strain genomic sequence and the CDC/HCV1 cDNA sequence can be determined by techniques known in the art. For example, they can be determined by a direct comparison of the sequence information of the polynucleotide from the putative HCV, and the HCV cDNA sequence(s) described herein. For example, also,

they can be determined by hybridization of the polynucleotides under conditions which form stable duplexes between homologous regions (for example, those  
5 which would be used prior to  $S_1$  digestion), followed by digestion with single stranded specific nuclease(s), followed by size determination of the digested fragments.

Because of the evolutionary relationship of the strains or isolates of HCV, putative HCV strains or  
10 isolates are identifiable by their homology at the polypeptide level. Generally, HCV strains or isolates are expected to be more than about 40% homologous, probably more than about 70% homologous, and even more probably  
15 more than about 80% homologous, and some may even be more than about 90% homologous at the polypeptide level. The techniques for determining amino acid sequence homology are known in the art. For example, the amino acid sequence may be determined directly and compared to the sequences provided herein. Alternatively the nucleotide  
20 sequence of the genomic material of the putative HCV may be determined (usually via a cDNA intermediate), the amino acid sequence encoded therein can be determined, and the corresponding regions compared.

As used herein, a polynucleotide "derived from"  
25 a designated sequence refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6 nucleotides, preferably at least about 8 nucleotides, more preferably at least about 10-12  
30 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding to a region of the designated nucleotide sequence. "Corresponding" means homologous to or complementary to the designated sequence. Preferably, the sequence of the region from which the polynucleotide is derived is homologous to or complementary to a sequence  
35 which is unique to an HCV genome. Whether or not a sequence is unique to the HCV genome can be determined by

techniques known to those of skill in the art. For example, the sequence can be compared to sequences in databanks, e.g., Genbank, to determine whether it is  
5 present in the uninfected host or other organisms. The sequence can also be compared to the known sequences of other viral agents, including those which are known to induce hepatitis, e.g., HAV, HBV, and HDV, and to other  
10 members of the Flaviviridae. The correspondence or non-correspondence of the derived sequence to other sequences can also be determined by hybridization under the appropriate stringency conditions. Hybridization techniques for determining the complementarity of nucleic acid  
15 sequences are known in the art, and are discussed infra. See also, for example, Maniatis et al. (1982). In addition, mismatches of duplex polynucleotides formed by hybridization can be determined by known techniques, including for example, digestion with a nuclease such as S1 that specifically digests single-stranded areas in  
20 duplex polynucleotides. Regions from which typical DNA sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

The derived polynucleotide is not necessarily  
25 physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated  
30 sequence may be modified in ways known in the art to be consistent with an intended use.

Similarly, a polypeptide or amino acid sequence  
"derived from" a designated nucleic acid sequence refers  
35 to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least

3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence.

A recombinant or derived polypeptide is not necessarily translated from a designated nucleic acid sequence, for example, the HCV cDNA sequences described herein, or from an HCV genome; it may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from mutated HCV. A recombinant or derived polypeptide may include one or more analogs of amino acids or unnatural amino acids in its sequence. Methods of inserting analogs of amino acids into a sequence are known in the art. It also may include one or more labels, which are known to those of skill in the art.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation which: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, or (3) does not occur in nature.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA, as well as double- and single stranded RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl



phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those  
5 containing pendant moieties, such as, for example proteins (including for e.g., nucleases, toxins, antibodies, signal peptides, poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing  
10 chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

The term "purified viral polynucleotide" refers to  
15 to an HCV genome or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90% of polypeptides with which the viral polynucleotide is naturally associated. Techniques for purifying viral  
20 polynucleotides from viral particles are known in the art, and include for example, disruption of the particle with a chaotropic agent, differential extraction and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation  
25 according to density.

The term "purified viral polypeptide" refers to an HCV polypeptide or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably  
30 less than about 90%, of cellular components with which the viral polypeptide is naturally associated. Techniques for purifying viral polypeptides are known in the art, and examples of these techniques are discussed infra. The term  
35 "purified viral polynucleotide" refers to an HCV genome or fragment thereof which is essentially free, i.e., contains less than about 20%, preferably less than about 50%, and

even more preferably less than about 70% of polypeptides with which the viral polynucleotide is naturally associated. Techniques for purifying viral polynucleotides from viral particles are known in the art, and include for example, disruption of the particle with a chaotropic agent, and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc. that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers.

The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

"Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are also present in the designated polypeptide(s), usually HCV proteins. Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art, and are also illustrated infra.

As used herein, "epitope" refers to an antigenic determinant of a polypeptide; an epitope could comprise 3 amino acids in a spatial conformation which is unique to

the epitope, generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 8-10 such amino acids. Methods of determining the spatial  
5 conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance.

A polypeptide is "immunologically reactive" with an antibody when it binds to an antibody due to antibody  
10 recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an  
15 epitope against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art.

As used herein, the term "immunogenic  
20 polypeptide" is a polypeptide that elicits a cellular and/or humoral response, whether alone or linked to a carrier in the presence or absence of an adjuvant.

The term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the  
25 product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the  
30 like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.  
35

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

"Treatment" as used herein refers to prophylaxis and/or therapy.

An "individual", as used herein, refers to vertebrates, particularly members of the mammalian species, and includes but is not limited to domestic animals, sports animals, and primates, including humans.

As used herein, the "sense strand" of a nucleic acid contains the sequence that has sequence homology to that of mRNA. The "anti-sense strand" contains a sequence which is complementary to that of the "sense strand".

As used herein, a "positive stranded genome" of a virus is one in which the genome, whether RNA or DNA, is single-stranded and which encodes a viral polypeptide(s). Examples of positive stranded RNA viruses include Togaviridae, Coronaviridae, Retroviridae, Picornaviridae, and Caliciviridae. Included also, are the Flaviviridae, which were formerly classified as Togaviridae. See Fields & Knipe (1986).

As used herein, "antibody-containing body component" refers to a component of an individual's body which is a source of the antibodies of interest. Antibody containing body components are known in the art, and include but are not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, white blood cells, and myelomas.

As used herein, "purified HCV" refers to a preparation of HCV which has been isolated from the cellular constituents with which the virus is normally associated, and from other types of viruses which may be present in the infected tissue. The techniques for isolating viruses are known to those of skill in the art, and include, for example, centrifugation and affinity chromatography; a method of preparing purified HCV is discussed infra.

The term "HCV particles" as used herein include entire virion as well as particles which are intermediates in virion formation. HCV particles generally have one or more HCV proteins associated with the HCV nucleic acid.

As used herein, the term "probe" refers to a polynucleotide which forms a hybrid structure with a sequence in a target region, due to complementarity of at least one sequence in the probe with a sequence in the target region. The probe, however, does not contain a sequence complementary to sequence(s) used to prime the polymerase chain reaction.

As used herein, the term "target region" refers to a region of the nucleic acid which is to be amplified and/or detected.

As used herein, the term "viral RNA", which includes HCV RNA, refers to RNA from the viral genome, fragments thereof, transcripts thereof, and mutant sequences derived therefrom.

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of in vitro cell culture constituents (including but not limited to conditioned medium resulting from the

growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

5    II. Description of the Invention

          The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitch & Sambrook, MOLECULAR CLONING; A LABORATORY MANUAL (1982); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

          The useful materials and processes of the present invention are made possible by the provision of a family of nucleotide sequences isolated from cDNA libraries which contain HCV cDNA sequences. These cDNA librar-

ies were derived from nucleic acid sequences present in the plasma of an HCV-infected chimpanzee. The construction of one of these libraries, the "c" library (ATCC No. 40394), was reported in EPO Pub. No. 318,216. Several of the clones containing HCV cDNA reported herein were obtained from the "c" library. Although other clones reported herein were obtained from other HCV cDNA libraries, the presence of clones containing the sequences in the "c" library was confirmed. As discussed in EPO Pub. No. 318,216, the family of HCV cDNA sequences isolated from the "c" library are not of human or chimpanzee origin, and show no significant homology to sequences contained within the HBV genome.

The availability of the HCV cDNAs described herein permits the construction of polynucleotide probes which are reagents useful for detecting viral polynucleotides in biological samples, including donated blood. For example, from the sequences it is possible to synthesize DNA oligomers of about 8-10 nucleotides, or larger, which are useful as hybridization probes to detect the presence of HCV RNA in, for example, donated blood, sera of subjects suspected of harboring the virus, or cell culture systems in which the virus is replicating. In addition, the cDNA sequences also allow the design and production of HCV specific polypeptides which are useful as diagnostic reagents for the presence of antibodies raised during HCV infection. Antibodies to purified polypeptides derived from the cDNAs may also be used to detect viral antigens in biological samples, including, for example, donated blood samples, sera from patients with NANBH, and in tissue culture systems being used for HCV replication. Moreover, the immunogenic polypeptides disclosed herein, which are encoded in portions of the ORF of HCV cDNA shown in Fig. 17, are also useful for HCV



screening, diagnosis, and treatment, and for raising antibodies which are also useful for these purposes.

5 In addition, the novel cDNA sequences described herein enable further characterization of the HCV genome. Polynucleotide probes and primers derived from these sequences may be used to amplify sequences present in cDNA libraries, and/or to screen cDNA libraries for additional  
10 overlapping cDNA sequences, which, in turn, may be used to obtain more overlapping sequences. As indicated infra. and in EPO Pub. No. 318,216, the genome of HCV appears to be RNA comprised primarily of a large open reading frame (ORF) which encodes a large polyprotein.

15 The HCV cDNA sequences provided herein, the polypeptides derived from these sequences, and the immunogenic polypeptides described herein, as well as antibodies directed against these polypeptides are also useful in the isolation and identification of the blood-borne NABV (BB-NANBV) agent(s). For example, antibodies  
20 directed against HCV epitopes contained in polypeptides derived from the cDNAs may be used in processes based upon affinity chromatography to isolate the virus. Alternatively, the antibodies may be used to identify  
25 viral particles isolated by other techniques. The viral antigens and the genomic material within the isolated viral particles may then be further characterized.

In addition to the above, the information provided infra allows the identification of additional HCV strains or isolates. The isolation and characterization  
30 of the additional HCV strains or isolates may be accomplished by isolating the nucleic acids from body components which contain viral particles and/or viral RNA, creating cDNA libraries using polynucleotide probes based on the HCV cDNA probes described infra., screening the  
35 libraries for clones containing HCV cDNA sequences described infra., and comparing the HCV cDNAs from the new

isolates with the cDNAs described infra. The polypeptides encoded therein, or in the viral genome, may be monitored for immunological cross-reactivity utilizing the polypeptides and antibodies described supra. Strains or isolates which fit within the parameters of HCV, as described in the Definitions section, supra., are readily identifiable. Other methods for identifying HCV strains will be obvious to those of skill in the art, based upon the information provided herein.

#### Isolation of the HCV cDNA Sequences

The novel HCV cDNA sequences described infra. extend the sequence of the cDNA to the HCV genome reported in EPO Pub. No. 318,216. The sequences which are present in clones b114a, 18g, ag30a, CA205a, CA290a, CA216a, p114a, CA167b, CA156e, CA84a, and CA59a lie upstream of the reported sequence, and when compiled, yield nucleotides nos. -319 to 1348 of the composite HCV cDNA sequence. (The negative number on a nucleotide indicates its distance upstream of the nucleotide which starts the putative initiator MET codon.) The sequences which are present in clones b5a and 16jh lie downstream of the reported sequence, and yield nucleotides nos. 8659 to 8866 of the composite sequence. The composite HCV cDNA sequence which includes the sequences in the aforementioned clones, is shown in Fig. 17.

The novel HCV cDNAs described herein were isolated from a number of HCV cDNA libraries, including the "c" library present in lambda gt11 (ATCC No. 40394). The HCV cDNA libraries were constructed using pooled serum from a chimpanzee with chronic HCV infection and containing a high titer of the virus, i.e., at least  $10^6$  chimp infectious doses/ml (CID/ml). The pooled serum was used to isolate viral particles; nucleic acids isolated from these particles was used as the template in the construc-

tion of cDNA libraries to the viral genome. The procedures for isolation of putative HCV particles and for constructing the "c" HCV cDNA library is described in EPO Pub. No. 318,216. Other methods for constructing HCV cDNA libraries are known in the art, and some of these methods are described infra., in the Examples. Isolation of the sequences was by screening the libraries using synthetic polynucleotide probes, the sequences of which were derived from the 5'-region and the 3'-region of the known HCV cDNA sequence. The description of the method to retrieve the cDNA sequences is mostly of historical interest. The resultant sequences (and their complements) are provided herein, and the sequences, or any portion thereof, could be prepared using synthetic methods, or by a combination of synthetic methods with retrieval of partial sequences using methods similar to those described herein.

#### Preparation of Viral Polypeptides and Fragments

The availability of HCV cDNA sequences, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding antigenically active regions of the polypeptide encoded in either strand. These antigenically active regions may be derived from coat or envelope antigens or from core antigens, or from antigens which are non-structural including, for example, polynucleotide binding proteins, polynucleotide polymerase(s), and other viral proteins required for the replication and/or assembly of the virus particle. Fragments encoding the desired polypeptides are derived from the cDNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors

which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published  
5 October 1, 1986. Vectors for the expression of fusion polypeptides of SOD and HCV polypeptides encoded in a number of HCV clones are described infra., in the Examples. Any desired portion of the HCV cDNA containing an open reading frame, in either sense strand, can be  
10 obtained as a recombinant polypeptide, such as a mature or fusion protein; alternatively, a polypeptide encoded in the cDNA can be provided by chemical synthesis.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not  
15 containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host systems are presently used in forming recombinant polypeptides, and a summary of some of the more common  
20 control systems and host cell lines is given infra. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt  
25 fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such polypeptides can be used as diagnostics, or those which give rise to neutral-  
30 izing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy. In addition, as discussed infra., antibodies to these polypeptides are useful for isolating and identifying HCV  
35 particles.

Preparation of Antigenic Polypeptides and Conjugation with Carrier

5       An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of HCV antigen. Accordingly, using the  
10       cDNAs of HCV as a basis, DNAs encoding short segments of HCV polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the  
15       synthesized polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

      A number of techniques for obtaining such linkage are known in the art, including the formation of  
20       disulfide linkages using N-succinimidyl-3-(2-pyridylthio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks  
25       a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are  
30       known. See, for example, Immun. Rev. (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic  
35       acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The

carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the  
5 rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

10 Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as  
15 latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet  
20 hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

In addition to full-length viral proteins, polypeptides comprising truncated HCV amino acid sequences encoding at least one viral epitope are useful im-  
25 munological reagents. For example, polypeptides comprising such truncated sequences can be used as reagents in an immunoassay. These polypeptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While these truncated sequences can be  
30 produced by various known treatments of native viral protein, it is generally preferred to make synthetic or recombinant polypeptides comprising an HCV sequence. Polypeptides comprising these truncated HCV sequences can  
35 be made up entirely of HCV sequences (one or more epitopes, either contiguous or noncontiguous), or HCV sequences and heterologous sequences in a fusion protein.

Useful heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the HCV epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, e.g., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783, the disclosures of which are incorporated herein by reference.

The size of polypeptides comprising the truncated HCV sequences can vary widely, the minimum size being a sequence of sufficient size to provide an HCV epitope, while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired HCV epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated HCV amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the HCV sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select HCV sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

Truncated HCV amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire viral protein sequence can be screened by preparing a series of short peptides that together span the entire protein sequence. An example of antigenic screening of the regions of the HCV polyprotein is shown infra. In addition, by starting with, for example, 100mer polypeptides, it would be routine to test each polypeptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified 100mer to map the epitope of interest. Screening such peptides in an

immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare oligopeptides comprising the identified regions for screening. Such a computer analysis of the HCV amino acid sequence is shown in Fig. 20, where the hydrophilic/hydrophobic character is displayed above the antigen index. The amino acids are numbered from the starting MET (position 1) as shown in Fig. 17. It is appreciated by those of skill in the art that such computer analysis of antigenicity does not always identify an epitope that actually exists, and can also incorrectly identify a region of the protein as containing an epitope.

Examples of HCV amino acid sequences that may be useful, which are expressed from expression vectors comprised of clones 5-1-1, 81, CA74a, 35f, 279a, C36, C33b, CA290a, C8f, C12f, 14c, 15e, C25c, C33c, C33f, 33g, C39c, C40b, CA167b are described infra. Other examples of HCV amino acid sequences that may be useful as described herein are set forth below. It is to be understood that these peptides do not necessarily precisely map one epitope, and may also contain HCV sequence that is not immunogenic. These non-immunogenic portions of the sequence can be defined as described above using conventional techniques and deleted from the described sequences. Further, additional truncated HCV amino acid sequences that comprise an epitope or are immunogenic can be identified as described above. The following sequences are given by amino acid number (i.e., "AAn") where n is the amino acid number as shown in Fig. 17:

AA1-AA25; AA1-AA50; AA1-AA84; AA9-AA177; AA1-AA10;  
AA5-AA20; AA20-AA25; AA35-AA45; AA50-AA100;  
AA40-AA90; AA45-AA65; AA65-AA75; AA80-90; AA99-AA120;  
AA95-AA110; AA105-AA120; AA100-AA150; AA150-AA200;  
AA155-AA170; AA190-AA210; AA200-AA250; AA220-AA240;



5 AA245-AA265; AA250-AA300; AA290-AA330; AA290-305;  
AA300-AA350; AA310-AA330; AA350-AA400; AA380-AA395;  
AA405-AA495; AA400-AA450; AA405-AA415; AA415-AA425;  
AA425-AA435; AA437-AA582; AA450-AA500; AA440-AA460;  
AA460-AA470; AA475-AA495; AA500-AA550; AA511-AA690;  
AA515-AA550; AA550-AA600; AA550-AA625; AA575-AA605;  
AA585-AA600; AA600-AA650; AA600-AA625; AA635-AA665;  
AA650-AA700; AA645-AA680; AA700-AA750; AA700-AA725;  
AA700-AA750; AA725-AA775; AA770-AA790; AA750-AA800;  
AA800-AA815; AA825-AA850; AA850-AA875; AA800-AA850;  
AA920-AA990; AA850-AA900; AA920-AA945; AA940-AA965;  
10 AA970-AA990; AA950-AA1000; AA1000-AA1060;  
AA1000-AA1025; AA1000-AA1050; AA1025-AA1040;  
AA1040-AA1055; AA1075-AA1175; AA1050-AA1200;  
AA1070-AA1100; AA1100-AA1130; AA1140-AA1165;  
AA1192-AA1457; AA1195-AA1250; AA1200-AA1225;  
AA1225-AA1250; AA1250-AA1300; AA1260-AA1310;  
AA1260-AA1280; AA1266-AA1428; AA1300-AA1350;  
AA1290-AA1310; AA1310-AA1340; AA1345-AA1405;  
15 AA1345-AA1365; AA1350-AA1400; AA1365-AA1380;  
AA1380-AA1405; AA1400-AA1450; AA1450-AA1500;  
AA1460-AA1475; AA1475-AA1515; AA1475-AA1500;  
AA1500-AA1550; AA1500-AA1515; AA1515-AA1550;  
AA1550-AA1600; AA1545-AA1560; AA1569-AA1931;  
AA1570-AA1590; AA1595-AA1610; AA1590-AA1650;  
AA1610-AA1645; AA1650-AA1690; AA1685-AA1770;  
20 AA1689-AA1805; AA1690-AA1720; AA1694-AA1735;  
AA1720-AA1745; AA1745-AA1770; AA1750-AA1800;  
AA1775-AA1810; AA1795-AA1850; AA1850-AA1900;  
AA1900-AA1950; AA1900-AA1920; AA1916-AA2021;  
AA1920-AA1940; AA1949-AA2124; AA1950-AA2000;  
AA1950-AA1985; AA1980-AA2000; AA2000-AA2050;  
AA2005-AA2025; AA2020-AA2045; AA2045-AA2100;  
25 AA2045-AA2070; AA2054-AA2223; AA2070-AA2100;  
AA2100-AA2150; AA2150-AA2200; AA2200-AA2250;  
AA2200-AA2325; AA2250-AA2330; AA2255-AA2270;  
AA2265-AA2280; AA2280-AA2290; AA2287-AA2385;  
AA2300-AA2350; AA2290-AA2310; AA2310-AA2330;  
AA2330-AA2350; AA2350-AA2400; AA2348-AA2464;  
AA2345-AA2415; AA2345-AA2375; AA2370-AA2410;  
AA2371-AA2502; AA2400-AA2450; AA2400-AA2425;  
30 AA2415-AA2450; AA2445-AA2500; AA2445-AA2475;  
AA2470-AA2490; AA2500-AA2550; AA2505-AA2540;  
AA2535-AA2560; AA2550-AA2600; AA2560-AA2580;  
AA2600-AA2650; AA2605-AA2620; AA2620-AA2650;  
AA2640-AA2660; AA2650-AA2700; AA2655-AA2670;  
AA2670-AA2700; AA2700-AA2750; AA2740-AA2760;  
AA2750-AA2800; AA2755-AA2780;  
AA2780-AA2830; AA2785-AA2810; AA2796-AA2886;  
35 AA2810-AA2825; AA2800-AA2850; AA2850-AA2900;  
AA2850-AA2865; AA2885-AA2905; AA2900-AA2950;

AA2910-AA2930; AA2925-AA2950; AA2945-end(C' terminal).

5 The above HCV amino acid sequences can be prepared as discrete peptides or incorporated into a larger polypeptide, and may find use as described herein. Additional polypeptides comprising truncated HCV sequences are described in the examples.

10 The observed relationship of the putative polyproteins of HCV and the Flaviviruses allows some prediction of the putative domains of the HCV "non-structural" (NS) proteins. The locations of the individual NS proteins in the putative Flavivirus precursor polyprotein are fairly well-known. Moreover, these  
15 also coincide with observed gross fluctuations in the hydrophobicity profile of the polyprotein. It is established that NS5 of Flaviviruses encodes the virion polymerase, and that NS1 corresponds with a complement fixation antigen which has been shown to be an effective  
20 vaccine in animals. Recently, it has been shown that a flaviviral protease function resides in NS3. Due to the observed similarities between HCV and the Flaviviruses, described infra., deductions concerning the approximate locations of the corresponding protein domains and func-  
25 tions in the HCV polyprotein are possible. The expression of polypeptides containing these domains in a variety of recombinant host cells, including, for example, bacteria, yeast, insect, and vertebrate cells, should give rise to important immunological reagents which can be used for  
30 diagnosis, detection, and vaccines.

Although the non-structural protein regions of the putative polyproteins of the HCV isolate described herein and of Flaviviruses appear to have some similarity,  
35 there is less similarity between the putative structural regions which are towards the N-terminus. In this region, there is a greater divergence in sequence, and in addi-

tion, the hydrophobic profile of the two regions show less similarity. This "divergence" begins in the N-terminal region of the putative NS1 domain in HCV, and extends to the presumed N-terminus. Nevertheless, it may still be possible to predict the approximate locations of the putative nucleocapsid (N-terminal basic domain) and E (generally hydrophobic) domains within the HCV polyprotein. In the Examples the predictions are based on the changes observed in the hydrophobic profile of the HCV polyprotein, and on a knowledge of the location and character of the flaviviral proteins. From these predictions it may be possible to identify approximate regions of the HCV polyprotein that could correspond with useful immunological reagents. For example, the E and NS1 proteins of Flaviviruses are known to have efficacy as protective vaccines. These regions, as well as some which are shown to be antigenic in the HCV isolate described herein, for example those within putative NS3, C, and NS5, etc., should also provide diagnostic reagents. Moreover, the location and expression of viral-encoded enzymes may also allow the evaluation of anti-viral enzyme inhibitors, i.e., for example, inhibitors which prevent enzyme activity by virtue of an interaction with the enzyme itself, or substances which may prevent expression of the enzyme, (for example, anti-sense RNA, or other drugs which interfere with expression).

#### Preparation of Hybrid Particle Immunogens Containing HCV Epitopes

The immunogenicity of the epitopes of HCV may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. Constructs wherein the NANBV epitope is linked directly to the particle-forming protein

coding sequences produce hybrids which are immunogenic with respect to the HCV epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include HCV sequences are immunogenic with respect to HCV and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in S. cerevisiae (Valenzuela et al. (1982)), as well as in, for example, mammalian cells (Valenzuela, P., et al. (1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1986. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an HCV epitope. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the HCV epitope.

#### Preparation of Vaccines

Vaccines may be prepared from one or more immunogenic polypeptides derived from HCV cDNA, including the cDNA sequences described in the Examples. The

observed homology between HCV and Flaviviruses provides information concerning the polypeptides which may be most effective as vaccines, as well as the regions of the genome in which they are encoded. The general structure of the Flavivirus genome is discussed in Rice et al (1986). The flavivirus genomic RNA is believed to be the only virus-specific mRNA species, and it is translated into the three viral structural proteins, i.e., C, M, and E, as well as two large nonstructural proteins, NS4 and NS5, and a complex set of smaller nonstructural proteins. It is known that major neutralizing epitopes for Flaviviruses reside in the E (envelope) protein (Roehrig (1986)). Thus, vaccines may be comprised of recombinant polypeptides containing epitopes of HCV E. These polypeptides may be expressed in bacteria, yeast, or mammalian cells, or alternatively may be isolated from viral preparations. It is also anticipated that the other structural proteins may also contain epitopes which give rise to protective anti-HCV antibodies. Thus, polypeptides containing the epitopes of E, C, and M may also be used, whether singly or in combination, in HCV vaccines.

In addition to the above, it has been shown that immunization with NS1 (nonstructural protein 1), results in protection against yellow fever (Schlesinger et al (1986)). This is true even though the immunization does not give rise to neutralizing antibodies. Thus, particularly since this protein appears to be highly conserved among Flaviviruses, it is likely that HCV NS1 will also be protective against HCV infection. Moreover, it also shows that nonstructural proteins may provide protection against viral pathogenicity, even if they do not cause the production of neutralizing antibodies.

The information provided in the Examples concerning the immunogenicity of the polypeptides

expressed from cloned HCV cDNAs which span the various regions of the HCV ORF also allows predictions concerning their use in vaccines.

5 In view of the above, multivalent vaccines against HCV may be comprised of one or more epitopes from one or more structural proteins, and/or one or more epitopes from one or more nonstructural proteins. These vaccines may be comprised of, for example, recombinant HCV  
10 polypeptides and/or polypeptides isolated from the virions. In particular, vaccines are contemplated comprising one or more of the following HCV proteins, or subunit antigens derived therefrom: E, NS1, C, NS2, NS3, NS4 and NS5. Particularly preferred are vaccines compris-  
15 ing E and/or NS1, or subunits thereof.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are  
20 prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic  
25 ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vac-  
30 cine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-  
35 acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-

isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components  
5 extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of  
10 antibodies directed against an immunogenic polypeptide containing an HCV antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered  
15 parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For  
20 suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include  
25 such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions  
30 take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable  
35 salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with

the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic  
5 bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

#### Dosage and Administration of Vaccines

10 The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally  
15 in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend  
20 on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by  
25 other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the  
30 need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic HCV antigen(s) may be administered in conjunction with other immunoregulatory agents, for example, im-  
35 mune globulins.



### Preparation of Antibodies Against HCV Epitopes

The immunogenic polypeptides prepared as described above are used to produce antibodies, both polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing an HCV epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an HCV epitope contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Alternatively, polyclonal antibodies may be isolated from a mammal which has been previously infected with HCV. An example of a method for purifying antibodies to HCV epitopes from serum from an infected individual, based upon affinity chromatography and utilizing a fusion polypeptide of SOD and a polypeptide encoded within cDNA clone 5-1-1, is presented in EPO Pub. No. 318,216.

Monoclonal antibodies directed against HCV epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against HCV epitopes can be screened for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against HCV epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotypic antibodies.

Anti-idiotypic antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985).

Techniques for raising anti-idiotypic antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotypic antibodies may also be useful for treatment and/or diagnosis of NANBH, as well as for an elucidation of the immunogenic regions of HCV antigens.

It would also be recognized by one of ordinary skill in the art that a variety of types of antibodies directed against HCV epitopes may be produced. As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one antibody combining site. An "antibody combining site" or "binding domain" is formed from the folding of variable domains of an antibody molecule(s) to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows an immunological reaction with the antigen. An antibody combining site may be formed from a heavy and/or a light chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, altered antibodies, univalent

antibodies, the Fab proteins, and single domain antibodies.

5 A "single domain antibody" (dAb) is an antibody which is comprised of an VH domain, which reacts immunologically with a designated antigen. A dAb does not contain a VL domain, but may contain other antigen binding domains known to exist in antibodies, for example, the kappa and lambda domains. Methods for preparing dAbs are  
10 known in the art. See, for example, Ward et al. (1989).

Antibodies may also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for  
15 their preparation are known in the art (see, e.g., U.S. Patent No. 4,816,467, which is incorporated herein by reference), and include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and  
20 heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of all the chains of a particular antibody are homologous with the chains found in one anti-  
25 body produced by the lymphocyte which produces that antibody in situ, or in vitro (for example, in hybridomas). Vertebrate antibodies typically include native antibodies, for example, purified polyclonal antibodies and monoclonal antibodies. Examples of the methods for the  
30 preparation of these antibodies are described infra.

"Hybrid antibodies" are antibodies wherein one pair of heavy and light chains is homologous to those in a first antibody, while the other pair of heavy and light chains is homologous to those in a different second antibody. Typically, each of these two pairs will bind different epitopes, particularly on different antigens. This  
35 results in the property of "divalence", i.e., the ability

to bind two antigens simultaneously. Such hybrids may also be formed using chimeric chains, as set forth below.

5 "Chimeric antibodies", are antibodies in which the heavy and/or light chains are fusion proteins. Typically the constant domain of the chains is from one particular species and/or class, and the variable domains are from a different species and/or class. Also included is any antibody in which either or both of the heavy or  
10 light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be differing classes, or different species of origin, and whether or not the fusion point is at the variable/constant boundary. Thus, it is  
15 possible to produce antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose  
20 constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

Another example is "altered antibodies", which refers to antibodies in which the naturally occurring  
25 amino acid sequence in a vertebrate antibody has been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a  
30 region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region may be made to  
35 alter antigen binding characteristics. The antibody may also be engineered to aid the specific delivery of a

molecule or substance to a specific cell or tissue site. The desired alterations may be made by known techniques in molecular biology, e.g., recombinant techniques, site  
5 directed mutagenesis, etc.

Yet another example are "univalent antibodies", which are aggregates comprised of a heavy chain/light chain dimer bound to the Fc (i.e., constant) region of a second heavy chain. This type of antibody escapes  
10 antigenic modulation. See, e.g., Glennie et al. (1982).

Included also within the definition of antibodies are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the  
15 sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion. "Fab" includes aggregates  
20 of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as F(ab)<sub>2</sub>), which are capable of selectively reacting with a designated antigen or antigen family. "Fab" antibodies may be divided into subsets analogous to  
25 those described above, i.e., "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing "Fab" fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

30 II.H. Diagnostic Oligonucleotide Probes and Kits

Using the disclosed portions of the isolated HCV cDNAs as a basis, oligomers of approximately 8 nucleotides or more can be prepared, either by excision or  
35 synthetically, which hybridize with the HCV genome and are useful in identification of the viral agent(s), further characterization of the viral genome(s), as well as in

detection of the virus(es) in diseased individuals. The probes for HCV polynucleotides (natural or derived) are a length which allows the detection of unique viral sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 20 nucleotides appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. Among useful probes, for example, are those derived from the newly isolated clones disclosed herein, as well as the various oligomers useful in probing cDNA libraries, set forth below. A complement to any unique portion of the HCV genome will be satisfactory. For use as probes, complete complementarity is desirable, though it may be unnecessary as the length of the fragment is increased.

For use of such probes as diagnostics, the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. The probes are then labeled. Suitable labels, and methods for labeling probes are known in the art, and include, for example, radioactive labels incorporated by nick translation or kinasing, biotin, fluorescent probes, and chemiluminescent probes. The nucleic acids extracted from the sample are then treated with the labeled probe under hybridization conditions of suitable stringencies, and polynucleotide duplexes containing the probe are detected.

The probes can be made completely complementary to the HCV genome. Therefore, usually high stringency conditions are desirable in order to prevent false

positives. However, conditions of high stringency should only be used if the probes are complementary to regions of the viral genome which lack heterogeneity. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time, and concentration of formamide. These factors are outlined in, for example, Maniatis, T. (1982).

Generally, it is expected that the HCV genome sequences will be present in serum of infected individuals at relatively low levels, i.e., at approximately  $10^2$ - $10^3$  chimp infectious doses (CID) per ml. This level may require that amplification techniques be used in hybridization assays. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT-tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT application 84/03520 and EPA124221 describe a DNA hybridization assay in which: (1) analyte is annealed to a single-stranded DNA probe that is complementary to an enzyme-labeled oligonucleotide; and (2) the resulting tailed duplex is hybridized to an enzyme-labeled oligonucleotide. EPA 204510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands. A particularly desirable technique may first involve amplification of the target HCV sequences in sera approximately 10,000 fold, i.e., to approximately  $10^6$  sequences/ml. This may be accomplished, for example, by the polymerase chain reactions (PCR) technique described which is by Saiki et al. (1986),

by Mullis, U.S. Patent No. 4,683,195, and by Mullis et al. U.S. Patent No. 4,683,202. The amplified sequence(s) may then be detected using a hybridization assay which is  
5 described in EP 317,077, published May 24, 1989. These hybridization assays, which should detect sequences at the level of  $10^6$ /ml, utilize nucleic acid multimers which bind to single-stranded analyte nucleic acid, and which also  
10 bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO 225,807, published June 16, 1987.

The probes can be packaged into diagnostic kits.  
15 Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled and the ingredients for labeling may be included in the kit in separate containers. The kit may also contain other suitably packaged reagents and materials needed for the  
20 particular hybridization protocol, for example, standards, as well as instructions for conducting the test.

#### Immunoassay and Diagnostic Kits

Both the polypeptides which react immuno-  
25 logically with serum containing HCV antibodies, for example, those detected by the antigenic screening method described infra. in the Examples, as well those derived from or encoded within the isolated clones described in the Examples, and composites thereof, and the antibodies  
30 raised against the HCV specific epitopes in these polypeptides, are useful in immunoassays to detect presence of HCV antibodies, or the presence of the virus and/or viral antigens, in biological samples. Design of the  
35 immunoassays is subject to a great deal of variation, and a variety of these are known in the art. For example, the immunoassay may utilize one viral epitope; alternatively,



the immunoassay may use a combination of viral epitopes derived from these sources; these epitopes may be derived from the same or from different viral polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides. It may use, for example, a monoclonal antibody directed towards a viral epitope(s), a combination of monoclonal antibodies directed towards epitopes of one viral antigen, monoclonal antibodies directed towards epitopes of different viral antigens, polyclonal antibodies directed towards the same viral antigen, or polyclonal antibodies directed towards different viral antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Some of the antigenic regions of the putative polyprotein have been mapped and identified by screening the antigenicity of bacterial expression products of HCV cDNAs which encode portions of the polyprotein. See the Examples. Other antigenic regions of HCV may be detected by expressing the portions of the HCV cDNAs in other expression systems, including yeast systems and cellular systems derived from insects and vertebrates. In addition, studies giving rise to an antigenicity index and hydrophobicity/hydrophilicity profile give rise to information concerning the probability of a region's antigenicity.

The studies on antigenic mapping by expression of HCV cDNAs showed that a number of clones containing

these cDNAs expressed polypeptides which were immunologically reactive with serum from individuals with NANBH. No single polypeptide was immunologically reactive with all sera. Five of these polypeptides were very immunogenic in that antibodies to the HCV epitopes in these polypeptides were detected in many different patient sera, although the overlap in detection was not complete. Thus, the results on the immunogenicity of the polypeptides encoded in the various clones suggest that efficient detection systems may include the use of panels of epitopes. The epitopes in the panel may be constructed into one or multiple polypeptides.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing HCV epitopes or antibodies directed against HCV epitopes in suitable containers, along with the remaining reagents and materials required for the conduct of the assay, as well as a suitable set of assay instructions.

Further Characterization of the HCV Genome, Virions, and Viral Antigens Using Probes Derived From cDNA to the Viral Genome

The HCV cDNA sequence information in the newly isolated clones described in the Examples may be used to gain further information on the sequence of the HCV genome, and for identification and isolation of the HCV agent, and thus will aid in its characterization including the nature of the genome, the structure of the viral particle, and the nature of the antigens of which it is composed. This information, in turn, can lead to additional polynucleotide probes, polypeptides derived from the HCV genome, and antibodies directed against HCV

epitopes which would be useful for the diagnosis and/or treatment of HCV caused NANBH.

5       The cDNA sequence information in the above-mentioned clones is useful for the design of probes for the isolation of additional cDNA sequences which are derived from as yet undefined regions of the HCV genome(s) from which the cDNAs in clones described herein and in EP 0,316,218 are derived. For example, labeled probes  
10       containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the 5'-termini or 3'-termini of the composite HCV cDNA sequence shown in Fig. 17 may be used to isolate overlapping cDNA sequences from HCV cDNA  
15       libraries. Alternatively, characterization of the genomic segments could be from the viral genome(s) isolated from purified HCV particles. Methods for purifying HCV particles and for detecting them during the purification procedure are described herein, *infra*. Procedures for  
20       isolating polynucleotide genomes from viral particles are known in the art, and one procedure which may be used is that described in EP 0,218,316. The isolated genomic segments could then be cloned and sequenced. An example of  
25       this technique, which utilizes amplification of the sequences to be cloned, is provided *infra*., and yielded clone 16jh.

      Methods for constructing cDNA libraries are known in the art, and are discussed *supra* and *infra*; a  
30       method for the construction of HCV cDNA libraries in lambda-gt11 is discussed in EPO Pub. No. 318,216. However, cDNA libraries which are useful for screening with nucleic acid probes may also be constructed in other vectors known in the art, for example, lambda-gt10 (Huynh  
35       et al. (1985)).

Screening for Anti-Viral Agents for HCV

5       The availability of cell culture and animal  
model systems for HCV makes it possible to screen for  
anti-viral agents which inhibit HCV replication, and  
particularly for those agents which preferentially allow  
cell growth and multiplication while inhibiting viral  
replication. These screening methods are known by those  
10 of skill in the art. Generally, the anti-viral agents are  
tested at a variety of concentrations, for their effect on  
preventing viral replication in cell culture systems which  
support viral replication, and then for an inhibition of  
infectivity or of viral pathogenicity (and a low level of  
15 toxicity) in an animal model system.

      The methods and compositions provided herein for  
detecting HCV antigens and HCV polynucleotides are useful  
for screening of anti-viral agents in that they provide an  
alternative, and perhaps more sensitive means, for detect-  
20 ing the agent's effect on viral replication than the cell  
plaque assay or ID<sub>50</sub> assay. For example, the HCV-  
polynucleotide probes described herein may be used to  
quantitate the amount of viral nucleic acid produced in a  
cell culture. This could be accomplished, for example, by  
hybridization or competition hybridization of the infected  
25 cell nucleic acids with a labeled HCV-polynucleotide  
probe. For example, also, anti-HCV antibodies may be used  
to identify and quantitate HCV antigen(s) in the cell  
culture utilizing the immunoassays described herein. In  
addition, since it may be desirable to quantitate HCV  
30 antigens in the infected cell culture by a competition  
assay, the polypeptides encoded within the HCV cDNAs  
described herein are useful in these competition assays.  
Generally, a recombinant HCV polypeptide derived from the  
HCV cDNA would be labeled, and the inhibition of binding  
35 of this labeled polypeptide to an HCV polypeptide due to  
the antigen produced in the cell culture system would be

monitored. Moreover, these techniques are particularly useful in cases where the HCV may be able to replicate in a cell line without causing cell death.

5           The anti-viral agents which may be tested for efficacy by these methods are known in the art, and include, for example, those which interact with virion components and/or cellular components which are necessary for the binding and/or replication of the virus. Typical  
10 anti-viral agents may include, for example, inhibitors of virion polymerase and/or protease(s) necessary for cleavage of the precursor polypeptides. Other anti-viral agents may include those which act with nucleic acids to prevent viral replication, for example, anti-sense  
15 polynucleotides, etc.

          Antisense polynucleotides molecules are comprised of a complementary nucleotide sequence which allows them to hybridize specifically to designated regions of genomes or RNAs. Antisense polynucleotides may  
20 include, for example, molecules that will block protein translation by binding to mRNA, or may be molecules which prevent replication of viral RNA by transcriptase. They may also include molecules which carry agents (non-covalently attached or covalently bound) which cause the  
25 viral RNA to be inactive by causing, for example, scissions in the viral RNA. They may also bind to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability, or chronicity. Antisense molecules which are to hybridize to HCV derived  
30 RNAs may be designed based upon the sequence information of the HCV cDNAs provided herein. The antiviral agents based upon anti-sense polynucleotides for HCV may be designed to bind with high specificity, to be of increased solubility, to be stable, and to have low toxicity.  
35 Hence, they may be delivered in specialized systems, for example, liposomes, or by gene therapy. In addition, they

may include analogs, attached proteins, substituted or altered bonding between bases, etc.

5 Other types of drugs may be based upon polynucleotides which "mimic" important control regions of the HCV genome, and which may be therapeutic due to their interactions with key components of the system responsible for viral infectivity or replication.

#### 10 General Methods

The general techniques used in extracting the genome from a virus, preparing and probing a cDNA library, sequencing clones, constructing expression vectors, trans-  
15 forming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general guide, the following sets forth some  
20 sources currently available for such procedures, and for materials useful in carrying them out.

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, E. coli  
25 coli is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid  
30 containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic  
35 control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al.

(1977)), the tryptophan (*trp*) promoter system (Goeddel et al. (1980)) and the lambda-derived  $P_L$  promoter and N gene ribosome binding site (Shimatake et al. (1981)) and the hybrid *tac* promoter (De Boer et al. (1983)) derived from sequences of the *trp* and *lac* UV5 promoters. The foregoing systems are particularly compatible with *E. coli*; if desired, other prokaryotic hosts such as strains of *Bacillus* or *Pseudomonas* may be used, with corresponding control sequences.

Eukaryotic hosts include yeast and mammalian cells in culture systems. *Saccharomyces cerevisiae* and *Saccharomyces carlsbergensis* are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983)), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968); Holland et al. (1978)), including the promoter for 3 phosphoglycerate kinase (Hitzeman (1980)). Terminators may also be included, such as those derived from the enolase gene (Holland (1981)). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are

not naturally associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985, all of which are assigned to the herein assignee, and are hereby incorporated herein by reference.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978)), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art. Vectors suitable for replication in mammalian cells may include viral replicons, or sequences which insure integration of the appropriate sequences encoding NANBV epitopes into the host genome.

Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. For example, transformation of the E. coli host cells with lambda-gt11 containing BB-NANBV sequences is discussed in the Example section, infra. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972); Maniatis (1982)). Yeast



transformation by direct uptake may be carried out using the method of Hinnen et al. (1978). Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), or the various known modifications thereof.

Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 microgram of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 microliters buffer solution by incubation of 1-2 hr at 37° C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560.

Sticky ended cleavage fragments may be blunt ended using *E. coli* DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

Ligation mixtures are transformed into suitable cloning hosts, such as E. coli, and successful transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984). If desired the synthetic strands may be labeled with  $^{32}\text{P}$  by treatment with polynucleotide kinase in the presence of  $^{32}\text{P}$ -ATP, using standard conditions for the reaction.

DNA sequences, including those isolated from cDNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase using, as a primer, a synthetic oligonucleotide complementary to the portion of the DNA to be modified, and having the desired modification included in its own sequence. The resulting double stranded DNA is transformed into a phage supporting host bacterium. Cultures of the transformed bacteria, which contain replications of each strand of the phage, are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and conditions which permit hybridization with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

DNA libraries may be probed using the procedure of Grunstein and Hogness (1975). Briefly, in this procedure, the DNA to be probed is immobilized on nitro-cellulose filters, denatured, and prehybridized with a

buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinyl pyrrolidone, and Ficoll, 50 mM Na Phosphate (pH 6.5), 0.1% SDS, and 100 micrograms/ml carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42°C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'-<sup>32</sup>P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

For routine vector constructions, ligation mixtures are transformed into E. coli strain HB101 or other suitable host, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared according to the method of Clewell et al. (1969), usually following chloramphenicol amplification (Clewell (1972)). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing. Sequencing may be by the dideoxy method of Sanger et al. (1977) as further described by Messing et al. (1981), or by the method of Maxam et al. (1980). Problems with band compression, which are sometimes observed in GC rich regions, were

overcome by use of T-deazoguanosine according to Barr et al. (1986).

5           The enzyme-linked immunosorbent assay (ELISA)  
can be used to measure either antigen or antibody  
concentrations. This method depends upon conjugation of  
an enzyme to either an antigen or an antibody, and uses  
the bound enzyme activity as a quantitative label. To  
10       measure antibody, the known antigen is fixed to a solid  
phase (e.g., a microplate or plastic cup), incubated with  
test serum dilutions, washed, incubated with anti-  
immunoglobulin labeled with an enzyme, and washed again.  
Enzymes suitable for labeling are known in the art, and  
15       include, for example, horseradish peroxidase. Enzyme  
activity bound to the solid phase is measured by adding  
the specific substrate, and determining product formation  
or substrate utilization colorimetrically. The enzyme  
activity bound is a direct function of the amount of anti-  
body bound.

20           To measure antigen, a known specific antibody is  
fixed to the solid phase, the test material containing  
antigen is added, after an incubation the solid phase is  
washed, and a second enzyme-labeled antibody is added.  
25       After washing, substrate is added, and enzyme activity is  
estimated colorimetrically, and related to antigen  
concentration.

#### Examples

30           Described below are examples of the present  
invention which are provided only for illustrative  
purposes, and not to limit the scope of the present inven-  
tion. In light of the present disclosure, numerous  
embodiments within the scope of the claims will be appar-  
35       ent to those of ordinary skill in the art.

Isolation and Sequence of Overlapping

HCV cDNA Clones 13i, 26j, CA59a, CA84a, CA156e and CA167b

5       The clones 13i, 26j, CA59a, CA84a, CA156e and CA167b were isolated from the lambda-gt11 library which contains HCV cDNA (ATCC No. 40394), the preparation of which is described in EPO Pub. No. 318,216 (published 31 May 1989), and WO 89/04669 (published 1 June 1989).

10       Screening of the library was with the probes described infra., using the method described in Huynh (1985). The frequencies with which positive clones appeared with the respective probes was about 1 in 50,000.

15       The isolation of clone 13i was accomplished using a synthetic probe derived from the sequence of clone 12f. The sequence of the probe was:

5' GAA CGT TGC GAT CTG GAA GAC AGG GAC AGG 3'.

20       The isolation of clone 26j was accomplished using a probe derived from the 5'-region of clone K9-1. The sequence of the probe was:

5' TAT CAG TTA TGC CAA CGG AAG CGG CCC CGA 3'.

25       The isolation procedures for clone 12f and for clone k9-1 (also called K9-1) are described in EPO Pub. No. 318,216, and their sequences are shown in Figs. 1 and 2, respectively. The HCV cDNA sequences of clones 13i and 26j, are shown in Figs. 4 and 5, respectively. Also shown  
30       are the amino acids encoded therein, as well as the overlap of clone 13i with clone 12f, and the overlap of clone 26j with clone 13i. The sequences for these clones confirmed the sequence of clone K9-1. Clone K9-1 had been  
35       isolated from a different HCV cDNA library (See EP 0,218,316).

5           Clone CA59a was isolated utilizing a probe based upon the sequence of the 5'-region of clone 26j. The sequence of this probe was:

5' CTG GTT AGC AGG GCT TTT CTA TCA CCA CAA 3'.

10           A probe derived from the sequence of clone CA59a was used to isolate clone CA84a. The sequence of the probe used for this isolation was:

5' AAG GTC CTG GTA GTG CTG CTG CTA TTT GCC 3'.

15           Clone CA156e was isolated using a probe derived from the sequence of clone CA84a. The sequence of the probe was:

5' ACT GGA CGA CGC AAG GTT GCA ATT GCT CTA 3'.

20           Clone CA167b was isolated using a probe derived from the sequence of clone CA 156e. The sequence of the probe was:

25           5' TTC GAC GTC ACA TCG ATC TGC TTG TCG GGA 3'.

30           The nucleotide sequences of the HCV cDNAs in clones CA59a, CA84a, CA156e, and CA167b, are shown Figs. 6, 7, 8, and 9, respectively. The amino acids encoded therein, as well as the overlap with the sequences of relevant clones, are also shown in the Figs.

#### Creation of "pi" HCV cDNA Library

35           A library of HCV cDNA, the "pi" library, was constructed from the same batch of infectious chimpanzee plasma used to construct the lambda-gt11 HCV cDNA library (ATCC No. 40394) described in EPO Pub. No. 318,216, and

utilizing essentially the same techniques. However, construction of the pi library utilized a primer-extension method, in which the primer for reverse transcriptase was based on the sequence of clone CA59A. The sequence of the primer was:

5' GGT GAC GTG GGT TTC 3'.

10

#### Isolation and Sequence of Clone pi14a

Screening of the "pi" HCV cDNA library described supra., with the probe used to isolate clone CA167b (See supra.) yielded clone pi14a. The clone contains about 800 base pairs of cDNA which overlaps clones CA167b, CA156e, CA84a and CA59a, which were isolated from the lambda gt-11 HCV cDNA library (ATCC No. 40394). In addition, pi14a also contains about 250 base pairs of DNA which are upstream of the HCV cDNA in clone CA167b.

20

#### Isolation and Sequence of Clones CA216a, CA290a and ag30a

Based on the sequence of clone CA167b a synthetic probe was made having the following sequence:

25

5' GGC TTT ACC ACG TCA CCA ATG ATT GCC CTA 3'

The above probe was used to screen the , which yielded clone CA216a, whose HCV sequences are shown in Fig. 10.

Another probe was made based on the sequence of clone CA216a having the following sequence:

30

5' TTT GGG TAA GGT CAT CGA TAC CCT TAC GTG 3'

Screening the lambda-gt11 library (ATCC No. 40394) with this probe yielded clone CA290a, the HCV sequences therein being shown in Fig. 11.

35

In a parallel approach, a primer-extension cDNA library was made using nucleic acid extracted from the same infectious plasma used in the original lambda-gt11 cDNA library described above. The primer used was based on the sequence of clones CA216a and CA290a:

5' GAA GCC GCA CGT AAG 3'

The cDNA library was made using methods similar to those described previously for libraries used in the isolation of clones pil4a and k9-1. The probe used to screen this library was based on the sequence of clone CA290a:

5' CCG GCG TAG GTC GCG CAA TTT GGG TAA 3'

Clone ag30a was isolated from the new library with the above probe, and contained about 670 basepairs of HCV sequence. See Fig. 12. Part of this sequence overlaps the HCV sequence of clones CA216a and CA290a. About 300 base-pairs of the ag30a sequence, however, is upstream of the sequence from clone CA290a. The non-overlapping sequence shows a start codon (\*) and stop codons that may indicate the start of the HCV ORF. Also indicated in Fig. 12 are putative small encoded peptides (#) which may play a role in regulating translation, as well as the putative first amino acid of the putative polypeptide (/), and downstream amino acids encoded therein.

#### Isolation and Sequence of Clone CA205a

Clone CA205a was isolated from the original lambda gt-11 library (ATCC No. 40394), using a synthetic probe derived from the HCV sequence in clone CA290a (Fig. 11). The sequence of the probe was:

5' TCA GAT CGT TGG TGG AGT TTA CTT GTT GCC 3'.



5 The sequence of the HCV cDNA in CA205a, shown in Fig. 13, overlaps with the cDNA sequences in both clones ag30a and CA290a. The overlap of the sequence with that of CA290a is shown by the dotted line above the sequence (the figure also shows the putative amino acids encoded in this fragment).

10 As observed from the HCV cDNA sequences in clones CA205a and ag30a, the putative HCV polyprotein appears to begin at the ATG start codon; the HCV sequences in both clones contain an in-frame, contiguous double stop codon (TGATAG) forty two nucleotides upstream from this ATG. The HCV ORF appears to begin after these stop  
15 codons, and to extend for at least 8907 nucleotides (See the composite HCV cDNA shown in Fig. 17).

#### Isolation and Sequence of Clone 18g

20 Based on the sequence of clone ag30a (See Fig. 12) and of an overlapping clone from the original lambda gt-11 library (ATCC No. 40394), CA230a, a synthetic probe was made having the following sequence:

5' CCA TAG TGG TCT GCG GAA CCG GTG AGT ACA 3'.

25 Screening of the original lambda-gt11 HCV cDNA library with the probe yielded clone 18g, the HCV cDNA sequence of which is shown in Fig. 14. Also shown in the figure are the overlap with clone ag30a, and putative polypeptides encoded within the HCV cDNA.

30 The cDNA in clone 18g (C18g or 18g) overlaps that in clones ag30a and CA205a, described supra. The sequence of C18g also contains the double stop codon region observed in clone ag30a. The polynucleotide region upstream of these stop codons presumably represents part  
35 of the 5'-region of the HCV genome, which may contain short ORFs, and which can be confirmed by direct sequenc-

ing of the purified HCV genome. These putative small encoded peptides may play a regulatory role in translation. The region of the HCV genome upstream of that  
5 represented by C18g can be isolated for sequence analysis using essentially the technique described in EPO Pub. No. 318,216 for isolating cDNA sequences upstream of the HCV cDNA sequence in clone 12f. Essentially, small synthetic  
10 oligonucleotide primers of reverse transcriptase, which are based upon the sequence of C18g, are synthesized and used to bind to the corresponding sequence in HCV genomic RNA. The primer sequences are proximal to the known 5'-terminal of C18g, but sufficiently downstream to allow the  
15 design of probe sequences upstream of the primer sequences. Known standard methods of priming and cloning are used. The resulting cDNA libraries are screened with sequences upstream of the priming sites (as deduced from the elucidated sequence of C18g). The HCV genomic RNA is  
20 obtained from either plasma or liver samples from individuals with NANBH. Since HCV appears to be a Flavivirus-like virus, the 5'-terminus of the genome may be modified with a "cap" structure. It is known that Flavivirus  
25 genomes contain 5'-terminal "cap" structures. (Yellow Fever virus, Rice et al. (1988); Dengue virus, Hahn et al (1988); Japanese Encephalitis Virus (1987)).

Isolation and Sequence of Clones from  
the beta-HCV cDNA library

30 Clones containing cDNA representative of the 3'-terminal region of the HCV genome were isolated from a cDNA library constructed from the original infectious chimpanzee plasma pool which was used for the creation of the HCV cDNA lambda-gt11 library (ATCC No. 40394),  
35 described in EPO Pub. No. 318,216. In order to create the DNA library, RNA extracted from the plasma was "tailed" with poly rA using poly (rA) polymerase, and cDNA was

synthesized using oligo(dT)<sub>12-18</sub> as a primer for reverse transcriptase. The resulting RNA:cDNA hybrid was digested with RNAase H, and converted to double stranded HCV cDNA. The resulting HCV cDNA was cloned into lambda-gt10, using essentially the technique described in Huynh (1985), yielding the beta (or b) HCV cDNA library. The procedures used were as follows.

An aliquot (12ml) of the plasma was treated with proteinase K, and extracted with an equal volume of phenol saturated with 0.05M Tris-Cl, pH 7.5, 0.05% (v/v) beta-mercaptoethanol, 0.1% (w/v) hydroxyquinolone, 1 mM EDTA. The resulting aqueous phase was re-extracted with the phenol mixture, followed by 3 extractions with a 1:1 mixture containing phenol and chloroform:isoamyl alcohol (24:1), followed by 2 extractions with a mixture of chloroform and isoamyl alcohol (1:1). Subsequent to adjustment of the aqueous phase to 200 mM with respect to NaCl, nucleic acids in the aqueous phase were precipitated overnight at -20°C, with 2.5 volumes of cold absolute ethanol. The precipitates were collected by centrifugation at 10,000 RPM for 40 min., washed with 70% ethanol containing 20 mM NaCl, and with 100% cold ethanol, dried for 5 min. in a dessicator, and dissolved in water.

The isolated nucleic acids from the infectious chimpanzee plasma pool were tailed with poly rA utilizing poly-A polymerase in the presence of human placenta ribonuclease inhibitor (HPRI) (purchased from Amersham Corp.), utilizing MS2 RNA as carrier. Isolated nucleic acids equivalent to that in 2 ml of plasma were incubated in a solution containing TMN (50 mM Tris HCl, pH 7.9, 10 mM MgCl<sub>2</sub>, 250 mM NaCl, 2.5 mM MnCl<sub>2</sub>, 2 mM dithiothreitol (DTT)), 40 micromolar alpha-[<sup>32</sup>P] ATP, 20 units HPRI (Amersham Corp.), and about 9 to 10 units of RNase free poly-A polymerase (BRL). Incubation was for 10 min. at 37°C, and the reactions were stopped with EDTA (final

concentration about 250 mM). The solution was extracted with an equal volume of phenol-chloroform, and with an equal volume of chloroform, and nucleic acids were precipitated overnight at -20°C with 2.5 volumes of ethanol in the presence of 200 mM NaCl.

#### Isolation of Clone b5a

The beta HCV cDNA library was screened by hybridization using a synthetic probe, which had a sequence based upon the HCV cDNA sequence in clone 15e. The isolation of clone 15e is described in EPO Pub. No. 318,216, and its sequence is shown in Fig. 3. The sequence of the synthetic probe was:

5' ATT GCG AGA TCT ACG GGG CCT GCT ACT CCA 3'.

Screening of the library yielded clone beta-5a (b5a), which contains an HCV cDNA region of approximately 1000 base pairs. The 5'-region of this cDNA overlaps clones 35f, 19g, 26g, and 15e (these clones are described supra). The region between the 3'-terminal poly-A sequence and the 3'-sequence which overlaps clone 15e, contains approximately 200 base pairs. This clone allows the identification of a region of the 3'-terminal sequence the HCV genome.

The sequence of b5a is contained within the sequence of the HCV cDNA in clone 16jh (described infra). Moreover, the sequence is also present in CC34a, isolated from the original lambda-gt11 library (ATCC No. 40394). (The original lambda-gt11 library is referred to herein as the "C" library).

Isolation and Sequence of Clones Generated by PCR  
Amplification of the 3'-Region of the HCV Genome

5 Multiple cDNA clones have been generated which  
contain nucleotide sequences derived from the 3'-region of  
the HCV genome. This was accomplished by amplifying a  
targeted region of the genome by a polymerase chain re-  
action technique described in Saiki et al. (1986), and in  
10 Saiki et al. (1988), which was modified as described  
below. The HCV RNA which was amplified was obtained from  
the original infectious chimpanzee plasma pool which was  
used for the creation of the HCV cDNA lambda-gt11 library  
(ATCC No. 40394) described in EPO Pub. No. 318,216.  
15 Isolation of the HCV RNA was as described supra. The  
isolated RNA was tailed at the 3'-end with ATP by E. coli  
poly-A polymerase as described in Sippel (1973), except  
that the nucleic acids isolated from chimp serum were  
substituted for the nucleic acid substrate. The tailed  
20 RNA was then reverse transcribed into cDNA by reverse  
transcriptase, using an oligo dT-primer adapter, es-  
sentially as described by Han (1987), except that the  
components and sequence of the primer-adapter were:

25	<u>Stuffer</u>	<u>NotI</u>	<u>SP6 Promoter</u>	<u>Primer</u>
	AATTC	GCGGCCGC	CATACGATTTAGGTGACACTATAGAA	T <sub>15</sub>

The resultant cDNA was subjected to amplification by PCR  
using two primers:

30	<u>Primer</u>	<u>Sequence</u>
	JH32 (30mer)	ATAGCGGCCGCCCTCGATTGCGAGATCTAC
	JH11 (20mer)	AATTCGGGCGGCCGCCATACGA

35 The JH32 primer contained 20 nucleotide sequences  
hybridizable to the 5'-end of the target region in the  
cDNA, with an estimated  $T_m$  of 66°C. The JH11 was derived

from a portion of the oligo dT-primer adapter; thus, it is specific to the 3'-end of the cDNA with a  $T_m$  of 64°C. Both primers were designed to have a recognition site for the restriction enzyme, NotI, at the 5'-end, for use in subsequent cloning of the amplified HCV cDNA.

The PCR reaction was carried out by suspending the cDNA and the primers in 100 microliters of reaction mixture containing the four deoxynucleoside triphosphates, buffer salts and metal ions, and a thermostable DNA polymerase isolated from Thermus aquaticus (Taq polymerase), which are in a Perkin Elmer Cetus PCR kit (N801-0043 or N801-0055). The PCR reaction was performed for 35 cycles in a Perkin Elmer Cetus DNA thermal cycler. Each cycle consisted of a 1.5 min denaturation step at 94°C, an annealing step at 60°C for 2 min, and a primer extension step at 72°C for 3 min. The PCR products were subjected to Southern blot analysis using a 30 nucleotide probe, JH34, the sequence of which was based upon that of the 3'-terminal region of clone 15e. The sequence of JH34 is:

5' CTT GAT CTA CCT CCA ATC ATT CAA AGA CTC 3'.

The PCR products detected by the HCV cDNA probe ranged in size from about 50 to about 400 base pairs.

In order to clone the amplified HCV cDNA, the PCR products were cleaved with NotI and size selected by polyacrylamide gel electrophoresis. DNA larger than 300 base pairs was cloned into the NotI site of pUC18S. The vector pUC18S is constructed by including a NotI polylinker cloned between the EcoRI and Sali sites of pUC18. The clones were screened for HCV cDNA using the JH34 probe. A number of positive clones were obtained and sequenced. The nucleotide sequence of the HCV cDNA insert in one of these clones, 16jh, and the amino acids encoded

therein, are shown in Fig. 15. A nucleotide heterogeneity, detected in the sequence of the HCV cDNA in clone 16jh as compared to another clone of this region, is indicated in the figure.

#### Compiled HCV cDNA Sequences

An HCV cDNA sequence has been compiled from a series of overlapping clones derived from the various HCV cDNA libraries described supra.. In this sequence, the compiled HCV cDNA sequence obtained from clones b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, and CA59a is upstream of the compiled HCV cDNA sequence published in EPO Pub. No. 318,216, which is shown in Fig. 16. The compiled HCV cDNA sequence obtained from clones b5a and 16jh downstream of the compiled HCV cDNA sequence published in EPO Pub. No. 318,216.

Fig. 17 shows the compiled HCV cDNA sequence derived from the above-described clones and the compiled HCV cDNA sequence published in EPO Pub. No. 318,216. The clones from which the sequence was derived are b114a, 18g, ag30a, CA205a, CA290a, CA216a, pi14a, CA167b, CA156e, CA84a, CA59a, K9-1 (also called k9-1), 26j, 13i, 12f, 14i, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, 33f, 33g, 39c, 35f, 19g, 26g, 15e, b5a, and 16jh. In the figure the three dashes above the sequence indicate the position of the putative initiator methionine codon.

Clone b114a was obtained using the cloning procedure described for clone b5a, supra., except that the probe was the synthetic probe used to detect clone 18g, supra. Clone b114a overlaps with clones 18g, ag30a, and CA205a, except that clone b114a contains an extra two nucleotides upstream of the sequence in clone 18g (i.e., 5'-CA). These extra two nucleotides have been included in the HCV genomic sequence shown in Fig. 17.

It should be noted that although several of the clones described supra. have been obtained from libraries other than the original HCV cDNA lambda-gt11 C library (ATCC No. 40394), these clones contain HCV cDNA sequences which overlap HCV cDNA sequences in the original library. Thus, essentially all of the HCV sequence is derivable from the original lambda-gt11 C library (ATCC No. 40394) which was used to isolate the first HCV cDNA clone (5-1-1). The isolation of clone 5-1-1 is described in EPO Pub. No. 318,216.

Purification of Fusion Polypeptide C100-3  
(Alternate method)

The fusion polypeptide, C100-3 (also called HCV c100-3 and alternatively, c100-3), is comprised of superoxide dismutase (SOD) at the N-terminus an in-frame C100 HCV polypeptide at the C-terminus. A method for preparing the polypeptide by expression in yeast, and differential extraction of the insoluble fraction of the extracted host yeast cells, is described in EPO Pub. No. 318,216. An alternative method for the preparation of this fusion polypeptide is described below. In this method the antigen is precipitated from the crude cell lysate with acetone; the acetone precipitated antigen is then subjected to ion-exchange chromatography, and further purified by gel filtration.

The fusion polypeptide, C100-3 (HCV c100-3), is expressed in yeast strain JSC 308 (ATCC No. 20879) transformed with pAB24C100-3 (ATCC No. 67976); the transformed yeast are grown under conditions which allow expression (i.e., by growth in YEP containing 1% glucose). (See EPO Pub. No. 318,216). A cell lysate is prepared by suspending the cells in Buffer A (20 mM Tris HCl, pH 8.0, 1 mM EDTA, 1 mM PMSF. The cells are broken by grinding with glass beads in a Dynomill type homogenizer or its



equivalent. The extent of cell breakage is monitored by counting cells under a microscope with phase optics.

5 Broken cells appear dark, while viable cells are light-colored. The percentage of broken cells is determined.

When the percentage of broken cells is approximately 90% or greater, the broken cell debris is separated from the glass beads by centrifugation, and the glass beads are washed with Buffer A. After combining the washes and homogenate, the insoluble material in the lysate is obtained by centrifugation. The material in the pellet is washed to remove soluble proteins by suspension in Buffer B (50 mM glycine, pH 12.0, 1 mM DTT, 500 mM NaCl), followed by Buffer C (50 mM glycine, pH 10.0, 1 mM DTT). The insoluble material is recovered by centrifugation, and solubilized by suspension in Buffer C containing SDS. The extract solution may be heated in the presence of beta-mercaptoethanol and concentrated by ultrafiltration. The HCV c100-3 in the extract is precipitated with cold acetone. If desired, the precipitate may be stored at temperatures at about or below -15°C.

Prior to ion exchange chromatography, the acetone precipitated material is recovered by centrifugation, and may be dried under nitrogen. The precipitate is suspended in Buffer D (50 mM glycine, pH 10.0, 1 mM DTT, 7 M urea), and centrifuged to pellet insoluble material. The supernatant material is applied to an anion exchange column previously equilibrated with Buffer D. Fractions are collected and analyzed by ultraviolet absorbance or gel electrophoresis on SDS polyacrylamide gels. Those fractions containing the HCV c100-3 polypeptide are pooled.

35 In order to purify the HCV c100-3 polypeptide by gel filtration, the pooled fractions from the ion-exchange column are heated in the presence of beta-mercaptoethanol

and SDS, and the eluate is concentrated by ultrafiltration. The concentrate is applied to a gel filtration column previously equilibrated with Buffer E (20 mM Tris HCl, pH 7.0, 1 mM DTT, 0.1% SDS). The presence of HCV c100-3 in the eluted fractions, as well as the presence of impurities, are determined by gel electrophoresis on polyacrylamide gels in the presence of SDS and visualization of the polypeptides. Those fractions containing purified HCV c100-3 are pooled. Fractions high in HCV c100-3 may be further purified by repeating the gel filtration process. If the removal of particulate material is desired, the HCV c100-3 containing material may be filtered through a 0.22 micron filter.

Expression and Antigenicity of Polypeptides  
Encoded in HCV cDNA

Polypeptides Expressed in E. coli

The polypeptides encoded in a number of HCV cDNAs which span the HCV genomic ORF were expressed in E. coli, and tested for their antigenicity using serum obtained from a variety of individuals with NANBH. The expression vectors containing the cloned HCV cDNAs were constructed from pSODcfl (Steimer et al. (1986). In order to be certain that a correct reading frame would be achieved, three separate expression vectors, pcflAB, pcflCD, and pcflEF were created by ligating either of three linkers, AB, CD, and EF to a BamHI-EcoRI fragment derived by digesting to completion the vector pSODcfl with EcoRI and BamHI, followed by treatment with alkaline phosphatase. The linkers were created from six oligomers, A, B, C, D, E, and F. Each oligomer was phosphorylated by treatment with kinase in the presence of ATP prior to annealing to its complementary oligomer. The sequences of the synthetic linkers were the following.

	<u>Name</u>	<u>DNA Sequence (5' to 3')</u>
5	A	GATC CTG AAT TCC TGA TAA
	B	GAC TTA AGG ACT ATT TTA A
	C	GATC CGA ATT CTG TGA TAA
10	D	GCT TAA GAC ACT ATT TTA A
	E	GATC CTG GAA TTC TGA TAA
	F	GAC CTT AAG ACT ATT TTA A

Each of the three linkers destroys the original EcoRI site, and creates a new EcoRI site within the linker, but within a different reading frame. Hence, the HCV cDNA EcoRI fragments isolated from the clones when inserted into the expression vector, were in three different reading frames.

The HCV cDNA fragments in the designated lambda-gt11 clones were excised by digestion with EcoRI; each fragment was inserted into pcflAB, pcflCD, and pcflEF. These expression constructs were then transformed into D1210 E. coli cells, the transformants were cloned, and recombinant bacteria from each clone were induced to express the fusion polypeptides by growing the bacteria in the presence of IPTG.

Expression products of the indicated HCV cDNAs were tested for antigenicity by direct immunological screening of the colonies, using a modification of the method described in Helfman et al. (1983). Briefly, as shown in Fig. 18, the bacteria were plated onto nitrocellulose filters overlaid on ampicillin plates to give approximately 1,000 colonies per filter. Colonies were replica plated onto nitrocellulose filters, and the replicas were regrown overnight in the presence of 2 mM

5 IPTG and ampicillin. The bacterial colonies were lysed by  
suspending the nitrocellulose filters for about 15 to 20  
min in an atmosphere saturated with  $\text{CHCl}_3$  vapor. Each  
10 filter then was placed in an individual 100 mm Petri dish  
containing 10 ml of 50 mM Tris HCl, pH 7.5, 150 mM NaCl, 5  
mM  $\text{MgCl}_2$ , 3% (w/v) BSA, 40 micrograms/ml lysozyme, and 0.1  
microgram/ml DNase. The plates were agitated gently for  
at least 8 hours at room temperature. The filters were  
15 rinsed in TBST (50 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.005%  
Tween 20). After incubation, the cell residues were  
rinsed and incubated in TBS (TBST without Tween) contain-  
ing 10% sheep serum; incubation was for 1 hour. The  
filters were then incubated with pretreated sera in TBS  
20 from individuals with NANBH, which included: 3  
chimpanzees; 8 patients with chronic NANBH whose sera were  
positive with respect to antibodies to HCV C100-3  
polypeptide (described in EPO Pub. No. 318,216, and  
supra.) (also called C100); 8 patients with chronic NANBH  
25 whose sera were negative for anti-C100 antibodies; a  
convalescent patient whose serum was negative for anti-  
C100 antibodies; and 6 patients with community acquired  
NANBH, including one whose sera was strongly positive with  
respect to anti-C100 antibodies, and one whose sera was  
30 marginally positive with respect to anti-C100 antibodies.  
The sera, diluted in TBS, was pretreated by preabsorption  
with hSOD. Incubation of the filters with the sera was  
for at least two hours. After incubation, the filters  
were washed two times for 30 min with TBST. Labeling of  
35 expressed proteins to which antibodies in the sera bound  
was accomplished by incubation for 2 hours with  $^{125}\text{I}$ -  
labeled sheep anti-human antibody. After washing, the  
filters were washed twice for 30 min with TBST, dried, and  
autoradiographed.

A number of clones (see infra.) expressed  
polypeptides containing HCV epitopes which were im-

munologically reactive with serum from individuals with NANBH. Five of these polypeptides were very immunogenic in that antibodies to HCV epitopes in these polypeptides were detected in many different patient sera. The clones encoding these polypeptides, and the location of the polypeptide in the putative HCV polyprotein (wherein the amino acid numbers begin with the putative initiator codon) are the following: clone 5-1-1, amino acids 1694-1735; clone C100, amino acids 1569-1931; clone 33c, amino acids 1192-1457; clone CA279a, amino acids 1-84; and clone CA290a amino acids 9-177. The location of the immunogenic polypeptides within the putative HCV polyprotein are shown immediately below.

15

20

25

30

35

5      Clones encoding polypeptides of proven reactivity  
          with sera from NANBH patients.

	<u>Clone</u>	<u>Location within the HCV polyprotein</u> (amino acid no. beginning with puta- tive initiator methionine)
10		
	CA279a	1-84
	CA74a	437-582
	13i	511-690
	CA290a	9-177
15	33c	1192-1457
	40b	1266-1428
	5-1-1	1694-1735
	81	1689-1805
	33b	1916-2021
20	25c	1949-2124
	14c	2054-2223
	8f	2200-3325
	33f	2287-2385
	33g	2348-2464
25	39c	2371-2502
	15e	2796-2886
	C100	1569-1931

30      The results on the immunogenicity of the  
          polypeptides encoded in the various clones examined sug-  
          gest efficient detection and immunization systems may  
          include panels of HCV polypeptides/epitopes.

35      Expression of HCV Epitopes in Yeast  
          Three different yeast expression vectors which  
          allow the insertion of HCV cDNA into three different read-

ing frames are constructed. The construction of one of the vectors, pAB24C100-3 is described in EPO Pub. No. 318,216. In the studies below, the HCV cDNA from the clones listed in supra. in the antigenicity mapping study using the E. coli expressed products are substituted for the C100 HCV cDNA. The construction of the other vectors replaces the adaptor described in the above E. coli studies with one of the following adaptors:

Adaptor 1

ATT TTG AAT TCC TAA TGA G  
AC TTA AGG ATT ACT CAG CT

Adaptor 2

AAT TTG GAA TTC TAA TGA G  
AC CTT AAG ATT ACT CAG CT.

The inserted HCV cDNA is expressed in yeast transformed with the vectors, using the expression conditions described supra. for the expression of the fusion polypeptide, C100-3. The resulting polypeptides are screened using the sera from individuals with NANBH, described supra. for the screening of immunogenic polypeptides encoded in HCV cDNAs expressed in E. coli.

Comparison of the Hydrophobic Profiles of HCV Polyproteins with West Nile Virus Polyprotein and with Dengue Virus NS1

The hydrophobicity profile of an HCV polyprotein segment was compared with that of a typical Flavivirus, West Nile virus. The polypeptide sequence of the West Nile virus polyprotein was deduced from the known polynucleotide sequences encoding the non-structural

proteins of that virus. The HCV polyprotein sequence was deduced from the sequence of overlapping cDNA clones. The profiles were determined using an antigen program which  
5 uses a window of 7 amino acid width (the amino acid in question, and 3 residues on each side) to report the average hydrophobicity about a given amino acid residue. The parameters giving the reactive hydrophobicity for each amino acid residue are from Kyte and Doolittle (1982).  
10 Fig. 19 shows the hydrophobic profiles of the two polyproteins; the areas corresponding to the non-structural proteins of West Nile virus, ns1 through ns5, are indicated in the figure. As seen in the figure, there is a general similarity in the profiles of the HCV  
15 polyprotein and the West Nile virus polyprotein.

The sequence of the amino acids encoded in the 5'-region of HCV cDNA shown in Fig. 16 has been compared with the corresponding region of one of the strains of Dengue virus, described supra., with respect to the  
20 profile of regions of hydrophobicity and hydrophilicity (data not shown). This comparison indicated that the polypeptides from HCV and Dengue encoded in this region, which corresponds to the region encoding NS1 (or a portion thereof), have a similar hydrophobic/hydrophilic profile.  
25

The similarity in hydrophobicity profiles, in combination with the previously identified homologies in the amino acid sequences of HCV and Dengue Flavivirus in EP 0,218,316 suggests that HCV is related to these members of the Flavivirus family.  
30

#### Characterization of the Putative Polypeptides Encoded Within the HCV ORF

The sequence of the HCV cDNA sense strand, shown in Fig. 17, was deduced from the overlapping HCV cDNAs in  
35 the various clones described in EPO Pub. No. 318,216 and those described supra. It may be deduced from the



sequence that the HCV genome contains primarily one long continuous ORF, which encodes a polyprotein. In the sequence, nucleotide number 1 corresponds to the first nucleotide of the initiator MET codon; minus numbers indicate that the nucleotides are that distance away in the 5'-direction (upstream), while positive numbers indicate that the nucleotides are that distance away in the 3'-direction (downstream). The composite sequence shows the "sense" strand of the HCV cDNA.

The amino acid sequence of the putative HCV polyprotein deduced from the HCV cDNA sense strand sequence is also shown in Fig. 17, where position 1 begins with the putative initiator methionine.

Possible protein domains of the encoded HCV polyprotein, as well as the approximate boundaries, are the following (the polypeptides identified within the parentheses are those which are encoded in the Flavivirus domain):

	<u>Putative Domain</u>	<u>Approximate Boundary</u> <u>(amino acid nos.)</u>
25	"C" (nucleocapsid protein)	1-120
	"E" (Virion envelope protein(s) and possibly matrix (M) proteins	120-400
30	"NS1" (complement fixation antigen?)	400-660
	"NS2" (unknown function)	660-1050
35	"NS3" (protease?)	1050-1640

"NS4" (unknown function) 1640-2000

"NS5" (polymerase) 2000-? end

5

It should be noted, however, that hydrophobicity profiles (described infra), indicate that HCV diverges from the Flavivirus model, particularly with respect to the region upstream of NS2. Moreover, the boundaries indicated are not intended to show firm demarcations between the putative polypeptides.

15

The Hydrophilic and Antigenic Profile of the Polypeptide

Profiles of the hydrophilicity/hydrophobicity and the antigenic index of the putative polyprotein encoded in the HCV cDNA sequence shown in Fig. 16 were determined by computer analysis. The program for hydrophilicity/hydrophobicity was as described supra. The antigenic index results from a computer program which relies on the following criteria: 1) surface probability, 2) prediction of alpha-helicity by two different methods; 3) prediction of beta-sheet regions by two different methods; 4) prediction of U-turns by two different methods; 5) hydrophilicity/hydrophobicity; and flexibility. The traces of the profiles generated by the computer analyses are shown in Fig. 20. In the hydrophilicity profile, deflection above the abscissa indicates hydrophilicity, and below the abscissa indicates hydrophobicity. The probability that a polypeptide region is antigenic is usually considered to increase when there

is a deflection upward from the abscissa in the hydrophilic and/or antigenic profile. It should be noted, however, that these profiles are not necessarily indicators of the strength of the immunogenicity of a polypeptide.

#### Identification of Co-linear Peptides in HCV and Flaviviruses

The amino acid sequence of the putative polyprotein encoded in the HCV cDNA sense strand was compared with the known amino acid sequences of several members of Flaviviruses. The comparison shows that homology is slight, but due to the regions in which it is found, it is probably significant. The conserved co-linear regions are shown in Fig. 21. The amino acid numbers listed below the sequences represent the number in the putative HCV polyprotein (See Fig. 17.)

The spacing of these conserved motifs is similar between the Flaviviruses and HCV, and implies that there is some similarity between HCV and these flaviviral agents.

The following listed materials are on deposit under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

35

	<u>lambda-gt11</u>	<u>ATCC No.</u>	<u>Deposit Date</u>
	HCV cDNA library	40394	1 Dec. 1987
5	clone 81	40388	17 Nov. 1987
	clone 91	40389	17 Nov. 1987
	clone 1-2	40390	17 Nov. 1987
	clone 5-1-1	40391	18 Nov. 1987
	clone 12f	40514	10 Nov. 1988
10	clone 35f	40511	10 Nov. 1988
	clone 15e	40513	10 Nov. 1988
	clone K9-1	40512	10 Nov. 1988
	JSC 308	20879	5 May 1988
	pS356	67683	29 April 1988

- 15 In addition, the following deposits were made on 11 May 1989.

	<u>Strain</u>	<u>Linkers</u>	<u>ATCC No.</u>
20	D1210 (Cf1/5-1-1)	EF	67967
	D1210 (Cf1/81)	EF	67968
	D1210 (Cf1/CA74a)	EF	67969
	D1210 (Cf1/35f)	AB	67970
	D1210 (Cf1/279a)	EF	67971
	D1210 (Cf1/C36)	CD	67972
25	D1210 (Cf1/13i)	AB	67973
	D1210 (Cf1/C33b)	EF	67974
	D1210 (Cf1/CA290a)	AB	67975
	HB101 (AB24/C100 #3R)		67976

- 30 The following derivatives of strain D1210 were deposited on 3 May 1989.

	<u>Strain Derivative</u>	<u>ATCC No.</u>
	pCF1CS/C8f	67956
	pCF1AB/C12f	67952
5	pCF1EF/14c	67949
	pCF1EF/15e	67954
	pCF1AB/C25c	67958
	pCF1EF/C33c	67953
	pCF1EF/C33f	67050
10	pCF1CD/33g	67951
	pCF1CD/C39c	67955
	pCF1EF/C40b	67957
	pCF1EF/CA167b	67959

15 The following strains were deposited on May 12, 1989.

	<u>Strain</u>	<u>ATCC No.</u>
	Lambda gt11(C35)	40603
	Lambda gt10(beta-5a)	40602
20	D1210 (C40b)	67980
	D1210 (M16)	67981

25 Upon allowance and issuance of this application as a  
United States Patent, all restriction on availability of  
these deposits will be irrevocably removed; and access to  
the designated deposits will be available during pendency  
of the above-named application to one determined by the  
Commissioner to be entitled thereto under 37 CFR 1.14 and  
30 35 USC 1.22. Moreover, the designated deposits will be  
maintained for a period of thirty (30) years from the date  
of deposit, or for five (5) years after the last request  
for the deposit; or for the enforceable life of the U.S.  
patent, whichever is longer. The deposited materials  
35 mentioned herein are intended for convenience only, and  
are not required to practice the present invention in view

of the descriptions herein, and in addition these materials are incorporated herein by reference.

5

#### Industrial Applicability

The invention, in the various manifestations disclosed herein, has many industrial uses, some of which are the following. The HCV cDNAs may be used for the design of probes for the detection of HCV nucleic acids in  
10 samples. The probes derived from the cDNAs may be used to detect HCV nucleic acids in, for example, chemical synthetic reactions. They may also be used in screening programs for anti-viral agents, to determine the effect of the agents in inhibiting viral replication in cell culture  
15 systems, and animal model systems. The HCV polynucleotide probes are also useful in detecting viral nucleic acids in humans, and thus, may serve as a basis for diagnosis of HCV infections in humans.

In addition to the above, the cDNAs provided  
20 herein provide information and a means for synthesizing polypeptides containing epitopes of HCV. These polypeptides are useful in detecting antibodies to HCV antigens. A series of immunoassays for HCV infection, based on recombinant polypeptides containing HCV epitopes  
25 are described herein, and will find commercial use in diagnosing HCV induced NANBH, in screening blood bank donors for HCV-caused infectious hepatitis, and also for detecting contaminated blood from infectious blood donors. The viral antigens will also have utility in monitoring  
30 the efficacy of anti-viral agents in animal model systems. In addition, the polypeptides derived from the HCV cDNAs disclosed herein will have utility as vaccines for treatment of HCV infections.

The polypeptides derived from the HCV cDNAs,  
35 besides the above stated uses, are also useful for raising anti-HCV antibodies. Thus, they may be used in anti-HCV

vaccines. However, the antibodies produced as a result of immunization with the HCV polypeptides are also useful in detecting the presence of viral antigens in samples. Thus, they may be used to assay the production of HCV polypeptides in chemical systems. The anti-HCV antibodies may also be used to monitor the efficacy of anti-viral agents in screening programs where these agents are tested in tissue culture systems. They may also be used for passive immunotherapy, and to diagnose HCV caused NANBH by allowing the detection of viral antigen(s) in both blood donors and recipients. Another important use for anti-HCV antibodies is in affinity chromatography for the purification of virus and viral polypeptides. The purified virus and viral polypeptide preparations may be used in vaccines. However, the purified virus may also be useful for the development of cell culture systems in which HCV replicates.

Antisense polynucleotides may be used as inhibitors of viral replication.

For convenience, the anti-HCV antibodies and HCV polypeptides, whether natural or recombinant, may be packaged into kits.

CLAIMS

- 5           1. A recombinant polynucleotide comprising a  
sequence derived from HCV cDNA, wherein the HCV cDNA is in  
clone 13i, or clone 26j, or clone 59a, or clone 84a, or  
clone CA156e, or clone 167b, or clone pil4a, or clone  
CA216a, or clone CA290a, or clone ag30a, or clone 205a, or  
10       clone 18g, or clone 16jh, or wherein the HCV cDNA is of a  
sequence indicated by nucleotide numbers -319 to 1348 or  
8659 to 8866 in Fig. 17.
- 15           2. A recombinant polynucleotide according to  
claim 1, encoding an epitope of HCV.
3. A recombinant vector comprising the  
polynucleotide of claim 1 or claim 2.
- 20           4. A host cell transformed with the vector of  
claim 3.
5. A recombinant expression system comprising  
an open reading frame (ORF) of DNA derived from the re-  
combinant polynucleotide of claim 1 or claim 2, wherein  
25       the ORF is operably linked to a control sequence compat-  
ible with a desired host.
6. A cell transformed with the recombinant  
expression system of claim 5.  
30
7. A polypeptide produced by the cell of claim  
6.
- 35           8. A purified polypeptide comprising an epitope  
encoded within HCV cDNA wherein the HCV cDNA is of a



sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17.

5  
9. An immunogenic polypeptide produced by a cell transformed with a recombinant expression vector comprising an ORF of DNA derived from HCV cDNA, wherein the HCV cDNA is comprised of a sequence derived from the HCV cDNA sequence in clone CA279a, or clone CA74a, or  
10 clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the ORF is operably linked to a control sequence compatible with a desired host.  
15

10. A peptide comprising an HCV epitope, wherein the peptide is of the formula

20  $AA_x-AA_y$ ,

wherein x and y designate amino acid numbers shown in Fig. 17, and wherein the peptide is selected from the group consisting of AA1-AA25, AA1-AA50, AA1-AA84, AA9-AA177, AA1-AA10, AA5-AA20, AA20-AA25, AA35-AA45, AA50-AA100,  
25 AA40-AA90, AA45-AA65, AA65-AA75, AA80-90, AA99-AA120, AA95-AA110, AA105-AA120, AA100-AA150, AA150-AA200, AA155-AA170, AA190-AA210, AA200-AA250, AA220-AA240, AA245-AA265, AA250-AA300, AA290-AA330, AA290-305, AA300-AA350, AA310-AA330, AA350-AA400, AA380-AA395,  
30 AA405-AA495, AA400-AA450, AA405-AA415, AA415-AA425, AA425-AA435, AA437-AA582, AA450-AA500, AA440-AA460, AA460-AA470, AA475-AA495, AA500-AA550, AA511-AA690, AA515-AA550, AA550-AA600, AA550-AA625, AA575-AA605, AA585-AA600, AA600-AA650, AA600-AA625, AA635-AA665, AA650-AA700,  
35 AA645-AA680, AA700-AA750, AA700-AA725, AA700-AA750, AA725-AA775, AA770-AA790, AA750-AA800, AA800-AA815,

AA825-AA850, AA850-AA875, AA800-AA850, AA920-AA990,  
AA850-AA900, AA920-AA945, AA940-AA965, AA970-AA990,  
5 AA950-AA1000, AA1000-AA1060, AA1000-AA1025, AA1000-AA1050,  
AA1025-AA1040, AA1040-AA1055, AA1075-AA1175,  
AA1050-AA1200, AA1070-AA1100, AA1100-AA1130,  
AA1140-AA1165, AA1192-AA1457, AA1195-AA1250,  
AA1200-AA1225, AA1225-AA1250, AA1250-AA1300,  
10 AA1260-AA1310, AA1260-AA1280, AA1266-AA1428,  
AA1300-AA1350, AA1290-AA1310, AA1310-AA1340, AA1345-  
AA1405, AA1345-AA1365, AA1350-AA1400, AA1365-AA1380,  
AA1380-AA1405, AA1400-AA1450, AA1450-AA1500,  
AA1460-AA1475, AA1475-AA1515, AA1475-AA1500,  
AA1500-AA1550, AA1500-AA1515, AA1515-AA1550,  
15 AA1550-AA1600, AA1545-AA1560, AA1569-AA1931,  
AA1570-AA1590, AA1595-AA1610, AA1590-AA1650,  
AA1610-AA1645, AA1650-AA1690, AA1685-AA1770,  
AA1689-AA1805, AA1690-AA1720, AA1694-AA1735,  
AA1720-AA1745, AA1745-AA1770, AA1750-AA1800,  
20 AA1775-AA1810, AA1795-AA1850, AA1850-AA1900,  
AA1900-AA1950, AA1900-AA1920, AA1916-AA2021,  
AA1920-AA1940, AA1949-AA2124, AA1950-AA2000,  
AA1950-AA1985, AA1980-AA2000, AA2000-AA2050,  
AA2005-AA2025, AA2020-AA2045, AA2045-AA2100,  
25 AA2045-AA2070, AA2054-AA2223, AA2070-AA2100,  
AA2100-AA2150, AA2150-AA2200, AA2200-AA2250,  
AA2200-AA2325, AA2250-AA2330, AA2255-AA2270,  
AA2265-AA2280, AA2280-AA2290, AA2287-AA2385,  
AA2300-AA2350, AA2290-AA2310, AA2310-AA2330,  
30 AA2330-AA2350, AA2350-AA2400, AA2348-AA2464,  
AA2345-AA2415, AA2345-AA2375, AA2370-AA2410,  
AA2371-AA2502, AA2400-AA2450, AA2400-AA2425,  
AA2415-AA2450, AA2445-AA2500, AA2445-AA2475,  
AA2470-AA2490, AA2500-AA2550, AA2505-AA2540,  
35 AA2535-AA2560, AA2550-AA2600, AA2560-AA2580,  
AA2600-AA2650, AA2605-AA2620, AA2620-AA2650,

AA2640-AA2660, AA2650-AA2700, AA2655-AA2670,  
AA2670-AA2700, AA2700-AA2750, AA2740-AA2760,  
AA2750-AA2800, AA2755-AA2780,  
5 AA2780-AA2830, AA2785-AA2810, AA2796-AA2886,  
AA2810-AA2825, AA2800-AA2850, AA2850-AA2900,  
AA2850-AA2865, AA2885-AA2905, AA2900-AA2950,  
AA2910-AA2930, AA2925-AA2950, AA2945-end(C' terminal).

10 11. A polypeptide comprised of the peptide of  
claim 10.

12. An immunogenic polypeptide attached to a  
solid substrate, wherein the polypeptide is according to  
15 claim 7, or claim 8, or claim 9, or claim 10, or claim 11,  
or wherein the polypeptide is comprised of an epitope  
encoded within HCV cDNA wherein the HCV cDNA is of a  
sequence indicated by nucleotide numbers -319 to 1348 or  
8659 to 8866 in Fig. 17.  
20

13. A monoclonal antibody directed against an  
epitope encoded in HCV cDNA, wherein the HCV cDNA is of a  
sequence indicated by nucleotide numbers -319 to 1348 or  
8659 to 8866 in Fig. 17, or is the sequence present in  
25 clone 13i, or clone 26j, or clone 59a, or clone 84a, or  
clone CA156e, or clone 167b, or clone pil4a, or clone  
CA216a, or clone CA290a, or clone ag30a, or clone 205a, or  
clone 18g, or clone 16jh.

30 14. A preparation of purified polyclonal anti-  
bodies directed against a polypeptide comprised of an  
epitope encoded within HCV cDNA, wherein the HCV cDNA is  
of a sequence indicated by nucleotide numbers -319 to 1348  
or 8659 to 8866 in Fig. 17, or is the sequence present in  
35 clone 13i, or clone 26j, or clone 59a, or clone 84a, or  
clone CA156e, or clone 167b, or clone pil4a, or clone

CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

5           15. A polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or from the complement of the HCV cDNA sequence.

10

16. A kit for analyzing samples for the presence of polynucleotides from HCV comprising a polynucleotide probe containing a nucleotide sequence of about 8 or more nucleotides, wherein the nucleotide  
15           sequence is derived from HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, wherein the polynucleotide probe is in a suitable container.

20

17. A kit for analyzing samples for the presence of an HCV antigen comprising an antibody which reacts immunologically with an HCV antigen, wherein the antigen contains an epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or  
25           8659 to 8866 in Fig. 17, or wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone p114a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

30

18. A kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide containing an HCV epitope encoded within HCV cDNA which is of a sequence indicated by nucleotide  
35           numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or

5 clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh.

10 19. A kit for analyzing samples for the presence of an HCV antibody comprising an antigenic polypeptide expressed from HCV cDNA in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, wherein the antigenic polypeptide is present in a suitable container.

15 20. A method for detecting HCV nucleic acids in a sample comprising:

20 (a) reacting nucleic acids of the sample with a polynucleotide probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA sequence is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, and wherein the reacting is under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample,

25 (b) detecting a polynucleotide duplex which contains the probe, formed in step (a).

30 21. An immunoassay for detecting an HCV antigen comprising:

35 (a) incubating a sample suspected of containing an HCV antigen with an antibody directed against an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone

CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and (b) detecting an antibody-antigen complex formed in step (a) which contains the antibody.

22. An immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing anti-HCV antibodies with an antigen polypeptide containing an epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pi14a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the incubating is under conditions which allow formation of an antigen-antibody complex; and

(b) detecting an antibody-antigen complex formed in step (a) which contains the antigen polypeptide.

23. An immunoassay for detecting antibodies directed against an HCV antigen comprising:

(a) incubating a sample suspected of containing anti-HCV antibodies with the polypeptide of claim 9, under conditions which allow formation of an antigen-antibody complex; and

(b) detecting an antibody-antigen complex formed in step (a) which contains the antigen polypeptide.

24. A vaccine for treatment of HCV infection comprising an immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or

8659 to 8866 in Fig. 17 or is the sequence present in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

25. A method for producing antibodies to HCV comprising administering to an individual an isolated immunogenic polypeptide containing an HCV epitope encoded in HCV cDNA, wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17, or is of the sequence present in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33C or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or clone 39c, or clone 15e, and wherein the immunogenic polypeptide is present in a pharmacologically effective dose in a pharmaceutically acceptable excipient.

26. An antisense polynucleotide derived from HCV cDNA, wherein the HCV cDNA is that shown in Fig. 17.

27. A method for preparing purified fusion polypeptide C100-3 comprising:

- (a) providing a crude cell lysate containing polypeptide C100-3,
- (b) treating the crude cell lysate with an amount of acetone which causes the polypeptide to precipitate,
- (c) isolating and solubilizing the precipitated material,
- (d) isolating the C100-3 polypeptide by anion exchange chromatography, and

(e) further isolating the C100-3 polypeptide of step (d) by gel filtration.

5

28. A method for preparing an HCV polypeptide comprising:

10 (a) providing a host cell transformed with a recombinant expression system comprising an open reading frame (ORF) of DNA derived from HCV cDNA, wherein the HCV cDNA is in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone p114a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 15 1348 or 8659 to 8866 in Fig. 17, wherein the ORF is operably linked to a control sequence compatible with a desired host; and

(b) incubating the host cell under conditions with allow expression of the HCV polypeptide. 20

29. A method for preparing an immunogenic HCV polypeptide comprising:

25 (a) providing a host cell transformed with a recombinant expression vector comprising an ORF of DNA derived from HCV cDNA, wherein the HCV cDNA is comprised of a sequence derived from the HCV cDNA sequence in clone CA279a, or clone CA74a, or clone 13i, or clone CA290a, or clone 33c, or clone 40b, or clone 33b, or clone 25c, or clone 14c, or clone 8f, or clone 33f, or clone 33g, or 30 clone 39c, or clone 15e, wherein the OEF is operably linked to a control sequence compatible with the desired host; and

(b) incubating the host cell under conditions which allow expression of the HCV polypeptide. 35

30. A method for preparing a host cell



transformed with a recombinant polynucleotide comprising a sequence of HCV cDNA derived from the HCV cDNA in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17 comprising:

- (a) providing a host cell capable of transformation;
- (b) providing the recombinant polynucleotide; and
- (c) incubating (a) with (b) under conditions which allow transformation of the host cell with the polynucleotide.

31. A method for preparing a recombinant polynucleotide comprised of a sequence of HCV cDNA derived from the HCV cDNA in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e, or clone 167b, or clone pil4a, or clone CA216a, or clone CA290a, or clone ag30a, or clone 205a, or clone 18g, or clone 16jh, or wherein the HCV cDNA is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17 comprising:

- (a) providing a host cell transformed with the recombinant polynucleotide; and
- (b) isolating said polynucleotide from said host cell.

32. A method for preparing blood free of HCV comprising:

- (a) providing a sample of blood suspected of containing HCV and anti-HCV antibodies;

- (b) providing an immunogenic polypeptide prepared according to claim 28 or 29;
- 5 (c) incubating the sample of (a) with the immunogenic polypeptide of (b) under conditions which allow the formation of antibody-HCV polypeptide complexes;
- (d) detecting the complexes formed in step (c); and
- 10 (e) saving the blood from which complexes were not detected in (d).

33. A method for preparing blood free of HCV comprising:

- 15 (a) providing nucleic acids from a sample of blood suspected of containing HCV polynucleotides;
- (b) providing a probe for HCV, wherein the probe is comprised of an HCV sequence derived from an HCV cDNA which is of a sequence indicated by nucleotide numbers -319 to 1348 or 8659 to 8866 in Fig. 17,
- 20 (c) reacting (a) with (b) under conditions which allow the formation of a polynucleotide duplex between the probe and the HCV nucleic acid from the sample;
- (d) detecting a polynucleotide which contains the probe, formed in step (c); and
- 25 (e) saving the blood from which complexes were not detected in (d).

34. A method for producing a hybridoma which produces anti-HCV monoclonal antibodies comprising:

- 30 (a) immunizing an individual with an immunogenic polypeptide containing an epitope encoded in HCV cDNA, wherein the HCV cDNA is HCV cDNA in clone 13i, or clone 26j, or clone 59a, or clone 84a, or clone CA156e,
- 35 or clone 167b, or clone p114a, or clone CA216a, or clone

CA290a, or clone ag30a, or clone 205a, or clone 18g, or  
clone 16jh, or wherein the HCV cDNA is of a sequence  
indicated by nucleotide numbers -319 to 1348 or 8659 to  
5 8866 in Fig. 17; or

(b) immunizing an individual with an  
immunogenic polypeptide prepared according to claim 29;

(c) immortalizing antibody producing cells from  
the immunized individual;

10 (d) selecting an immortal cell which produces  
antibodies which react with an HCV epitope in the  
immunogenic polypeptide of (a) or (b); and

(e) growing said immortal cell.

15

20

25

30

35

## FIG. 1 Translation of DNA 12f

IlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsn  
1 CCATATTTAAAATCAGGATGTACGTGGGAGGGGTGGAACACAGGCTGGAAGCTGCCTGCA  
GGTATAAATTTTAGTCCTACATGCACCCCTCCCAGCTTGTGTCCGACCTTCGACGGACGT

TrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeu  
61 ACTGGACGCGGGGCGAACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGT  
TGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCA

LeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeu  
121 TACTGCTGACCACTACACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTACCAGCCT  
ATGACGACTGGTGATGTGTCACCGTCCAGGAGGGCACAAGGAAGTGTGGGATGGTCCGA

SerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGlnTyrLeuTyrGlyVal  
181 TGTCCACCGGCTCATCCACCTCCACCAGAACATTGTGGACGTGCAGTACTTGTACGGGG  
ACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAACACCTGCACGTCATGAACATGCCCC

-----  
GlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValValLeuLeuPheLeuLeu  
241 TGGGGTCAAGCATCGCGTCCTGGGCCATTAAGTGGGAGTACGTCGTTCTCCTGTTCCCTC  
ACCCAGTTCGTAGCGCAGGACCCGGTAATTCACCTCATGCAGCAAGAGGACAAGGAAG

-----  
LeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGlu  
301 TGCTTGACAGACGCGCGCTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGG  
ACGAACGTCTGCGCGCGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTGCGC

-----  
AlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeu  
361 AGGCGGCTTTGGAGAACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTC  
TCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAG

-----  
Val  
421 TTGTATC  
AACATAG

SUBSTITUTE SHEET

# SUBSTITUTE SHEET

3 / 40

ValCysSerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsn  
841 GCGTCTGCTCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAAGCGGCTTTGGAGA  
CGCAGACGAGGACGAACACCTACTACGATGAGTATAGGGTTCGCCTTCGCCGAAACCTCT

LeuValIleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuVal  
901 ACCTCGTAATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCG  
TGGAGCATTATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGC

PhePheCysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPhe  
961 TGTTCTTCTGCTTTGCATGGTATCTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCT  
ACAAGAAGACGAAACGTACCATAGACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGA

TyrGlyMetTrpProLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeu  
1021 TCTACGGGATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCGAGCGGGCGTACGCCG  
AGATGCCCTACACCGGAGAGGAGGACGAGGACAACCGCAACGGGGTTCGCCCGCATGCCGC

AspThrGluValAlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThr  
1081 TGGACACGGAGGTGGCCGCGTCTGTTGGCGGTGTTGTTCTCGTCGGGTGATGGCGCTAA  
ACCTGTGCCTCCACCGCGCAGCACACCGCCACAACAAGAGCAGCCCACTACCGCGATT

LeuSerProTyrTyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeu  
1141 CTCTGTCACCATATTACAAGCGCTATATCAGCTGGTGCTTGTGGTGGCTTCAGTATTTTC  
GAGACAGTGGTATAATGTTTCGGATATAGTCGACCACGAACACCACCGAAGTCATAAAAG

ThrArgValGluAlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArg  
1201 TGACCAGAGTGGAAGCGCAACTGCACGTGTGGATTCCCCCCTCAACGTCCGAGGGGGGC  
ACTGGTCTCACCTTCGCGTTGACGTGCACACCTAAGGGGGGAGTTGCAGGCTCCCCCG

AspAlaValIleLeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLys  
1261 GCGACGCTGTCATCTTACTCATGTGTGCTGTACACCCGACTCTGGTATTTGACATCACCA  
CGCTGCCAGTAGAATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGT

LeuLeuLeuAlaValPheGlyProLeuTrpIleLeuGlnAla  
1321 AATTGCTGCTGGCGTCTTCGGACCCCTTTGGATTCTTCAAGCCAG  
TTAACGACGACCGGCAGAAGCCTGGGGAAACCTAAGAAGTTCGGTC

FIG. 2-2

SUBSTITUTE SHEET

FIG. 3

1 GlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAla  
 CGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCCTACAACCCCCCTCGCGAGAGC  
 GCCGCGACCTTTCTCCAGATGATGGAGTGGGCACTGGGATGTTGGGGGGAGCGCTCTCG  
  
 61 AlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPhe  
 TCGTGAGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATGTT  
 ACGCACCTCTGTCGTTCTGTGTGAGGTCAAGTAAGGACCGATCCGTTGTATTAGTACAA  
  
 121 AlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAla  
 TGCCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTTCTTTAGCGTCTTATAGC  
 ACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCG  
  
 181 ArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGlu  
 CAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCTGCTACTCCATAGA  
 GTCCCTGGTCGAACCTGTCCGGGAGCTAACGCTCTAGATGCCCCGACGATGAGGTATCT  
  
 241 ProLeuAspLeuProProIleIleGlnArgLeu  
 ACCACTTGATCTACCTCCAATCATTCAAAGACTC  
 TGGTGAAC TAGATGGAGGTTAGTAAGTTTCTGAG

FIG. 5

Translation of DNA 26j

1 LeuPheTyrHisHisLysPheAsnSerSerGlyCysProGluArgLeuAlaSerCysArg  
 GCTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCTGAGAGGCTAGCCAGCTGCCG  
 CGAAAAGATAGTGGTGTTCAGTTGAGAAGTCCGACAGGACTCTCCGATCGGTGACGGC  
  
 61 ProLeuThrAspPheAspGlnGlyTrpGlyProIleSerTyrAlaAsnGlySerGlyPro  
 ACCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTATGCCAACGGAAGCGGCC  
 TGGGGAATGGCTAAACTGGTCCCGACCCGGGATAGTCAATACGGTTGCCTTCGCCGGG  
  
 121 AspGlnArgProTyrCysTrpHisTyrProProLysProCysGlyIleValProAlaLys  
 CGACCAGCGCCCCTACTGCTGGCACTACCCCCAAAACCTTGCGGTATTGTGCCCGCGAA  
 GCTGGTCGCGGGGATGACGACCGTGATGGGGGGTTTGGAACGCCATAACACGGGCGCTT  
  
 181 SerValCysGlyProValTyrCysPheThrProSerProValValVal  
 GAGTGTGTGTGGTCCGGTATATTGCTTCACTCCAGCCCCGTGGTGGTGGG  
 CTCACACACACCAGGCCATATAACGAAGTGAGGGTCGGGGCACCACCACC  
  
 ---Overlap with 131---

SUBSTITUTE SHEET

## FIG. 4

Translation of DNA 13i

1 ProSerProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGly  
 CTCCCAGCCCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCTACCTACAGCTGGG  
 GAGGGTCGGGGCACCACCACCTTGCTGGCTGTCCAGCCCCGCGGGATGGATGTCGACCC

61 GluAsnAspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPhe  
 GTGAAAATGATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGT  
 CACTTTTACTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCA

121 GlyCysThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysVal  
 TCGGTTGTACCTGGATGAACTCAACTGGATTACCAAAGTGTGCGGAGCGCTCCTTGTG  
 AGCCAACATGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACAC

181 IleGlyGlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisPro  
 TCATCGGAGGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATC  
 AGTAGCCTCCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTTCGTAG

241 AspAlaThrTyrSerArgCysGlySerGlyProTrpLeuThrProArgCysLeuValAsp  
 CGGACGCCACATACTCTCGGTGCGGCTCCGGTCCCTGGCTCACACCCAGGTGCCTGGTCTG  
 GCCTGCGGTGTATGAGAGCCACGCCGAGGCCAGGGACCGAGTGTGGGTCCACGGACCAGC

301 TyrProTyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArg  
 ACTACCCGTATAGGCTTTGGCATTATCCTTGTACCATCAACTACCCATATTTAAAATCA  
 TGATGGGCATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAATTTTAGT

361 MetTyrValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGlu  
 GGATGTACGTGGGAGGGGTCGAGCACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCG  
 CCTACATGCACCCTCCCCAGCTCGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGC

-----Overlap with 12f-----

421 ArgCysAspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThr  
 AACGTTGCGATCTGGAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTA  
 TTGCAACGCTAGACCTTCTGTCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGAT

481 GlnTrpGlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeu  
 CACAGTGGCAGGTCCTCCCGTGTTCCTTCACAACCCTGCCAGCCTTGTCACCGGCCTCA  
 GTGTCACCGTCCAGGAGGGCACAAGGAAGTGTGGGACGGTCGGAACAGGTGGCCGGAGT

SUBSTITUTE SHEET



## FIG. 6

## Translation of DNA CA59a

1 LeuValMetAlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIleAlaGlyAla  
TTGGTAATGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCT  
AACCATTACCGAGTCGACGAGGCCTAGGGTGTTCCGGTAGAACCTGTACTAGCGACCACGA

61 HisTrpGlyValLeuAlaGlyIleAlaTyrPheSerMetValGlyAsnTrpAlaLysVal  
CACTGGGGAGTCCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTC  
GTGACCCCTCAGGACCGCCCGTATCGCATAAAGAGGTACCACCCCTTGACCCGCTTCCAG

121 LeuValValLeuLeuLeuPheAlaGlyValAspAlaGluThrHisValThrGlyGlySer  
CTGGTAGTGCTGCTGCTATTGCGGGCGTCGACGCGGAAACCCACGTCACCGGGGGAAGT  
GACCATCACGACGACGATAAACGGCCGCGAGCTGCGCCTTTGGGTGCAGTGGCCCCCTTCA

181 AlaGlyHisThrValSerGlyPheValSerLeuLeuAlaProGlyAlaLysGlnAsnVal  
GCCGGCCACACTGTGTCTGGATTTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTC  
CGGCCGGTGTGACACAGACCTAAACAATCGGAGGAGCGTGGTCCGCGGTTCGTCTTGACG

241 GlnLeuIleAsnThrAsnGlySerTrpHisLeuAsnSerThrAlaLeuAsnCysAsnAsp  
CAGCTGATCAACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAAGTCAATGAT  
GTCGACTAGTTGTGGTTGCCGTCAACCGTGGAGTTATCGTGCCGGGACTTGACGTTACTA

301 SerLeuAsnThrGlyTrpLeuAlaGlyLeuPheTyrHisHisLysPheAsnSerSerGly  
AGCCTCAACACCGGCTGGTTGGCAGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGC  
TCGGAGTTGTGGCCGACCAACCGTCCCGAAAAGATAGTGGTGTTCAGTTGAGAAGTCCG  
-----Overlap with 26j-----

-----Overlap with K9-1-----  
361 CysProGluArgLeuAlaSerCysArgPro  
TGTCCTGAGAGGCTAGCCAGCTGCCGACCCC  
ACAGGACTCTCCGATCGGTCGACGGCTGGGG  
-----

SUBSTITUTE SHEET

7 / 40

## Translation of DNA CA84a

FIG. 7

```

      GlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMetAlaTrpAsp
1  CGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCGCATGGCATGGG
   GCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGCCAGTGGCGTACCGTACCC

      MetMetMetAsnTrpSerProThrThrAlaLeuValMetAlaGlnLeuLeuArgIlePro
61 ATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAATGGCTCAGCTGCTCCGGATCC
   TATACTACTACTTGACCAGGGGATGCTGCCGCAACCATTACCGAGTCGACGAGGCCTAGG

      GlnAlaIleLeuAspMetIleAlaGlyAlaHisTrpGlyValLeuAlaGlyIleAlaTyr
121 CACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGGAGTCTTGGCGGGCATAGCGT
   GTGTTCCGTAGAACCCTGTACTAGCGACCACGAGTGACCCCTCAGGACCGCCCGTATCGCA

-----Overlap with CA59a-----
      PheSerMetValGlyAsnTrpAlaLysValLeuValValLeuLeuPheAlaGlyVal
181 ATTTCTCCATGGTGGGAACTGGGCGAAGGTCCTGGTAGTGCTGCTGCTATTTGCCGGCG
   TAAAGAGGTACCACCCCTTGACCCGCTTCCAGGACCATCACGACGACGATAAACGGCCGC

      AspAlaGluThrHisValThrGly
241 TCGACGCGGAAACCCACGTCACCGGGG
   AGCTGCGCCTTTGGGTGCAGTGGCCCC

```

## Translation of DNA CA156e

FIG. 8

```

      CysTrpValAlaMetThrProThrValAlaThrArgAspGlyLysLeuProAlaThrGln
1  GTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACCTCCCCGCGACGCA
   CACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGGCGCTGCGT

      LeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCysSerAlaLeuTyrVal
61 GCTTCGACGTCACATCGATCTGCTTGTGCGGGAGCGCCACCCTCTGTTCCGGCCCTCTACGT
   CGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACAAGCCGGGAGATGCA

      GlyAspLeuCysGlySerValPheLeuValGlyGlnLeuPheThrPheSerProArgArg
121 GGGGGACCTATGCGGGTCTGTCTTTCTTGTGCGGCCAACTGTTACCTTCTCTCCCAGGCG
   CCCCTGGATACGCCCAGACAGAAAGAACAGCCGGTTGACAAGTGGAAGAGAGGGTCCGC

      HisTrpThrThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArg
181 CCACTGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCG
   GGTGACCTGCTGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGCCAGTGGC

-----Overlap with CA84a-----
      MetAlaTrpAspMetMetMetAsnTrpSerProThrThrAlaLeuValValAlaGlnLeu
241 CATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTGGTAGTGGCTCAGCT
   GTACCGTACCCTATACTACTACTTGACCAGGGGATGCTGCCGCAACCATCACCGAGTCGA

      LeuArgIleProGlnAla
301 GCTCCGGATCCCACAAGCC
   CGAGGCCTAGGGTGTTCGG

```

SUBSTITUTE SHEET

## FIG. 9

## Translation of DNA CA167b

SerThrGlyLeuTyrHisValThrAsnAspCysProAsnSerSerIleValTyrGluAla  
1 CTCCACGGGGCTTTACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGC  
GAGGTGCCCCGAAATGGTGCAGTGGTTACTAACGGGATTGAGCTCATAACACATGCTCCG

AlaAspAlaIleLeuHisThrProGlyCysValProCysValArgGluGlyAsnAlaSer  
61 GGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCAACGCCTC  
CCGGCTACGGTAGGACGTGTGAGGCCCCACGCAGGGAACGCAAGCACTCCCGTTGCGGAG

-----  
ArgCysTrpValAlaMetThrProThrValAlaThrArgAspGlyLysLeuProAlaThr  
121 GAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAAC TCCCCGCGAC  
CTCCACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGGCGCTG

-----Overlap with CA156e-----  
GlnLeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCysSerAlaLeuTyr  
181 GCAGCTTCGACGTCACATCGATCTGCTTGTCTGGGAGCGCTACCCTCTGTTCGGCCCTCTA  
CGTCGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGATGGGAGACAAGCCGGGAGAT

-----  
ValGlyAspLeuCysGlySerValPheLeu  
241 CGTGGGGGACTTGTGCGGGTCTGTCTTTCTTG  
GCACCCCTGAACACGCCCAGACAGAAAGAAC

SUBSTITUTE SHEET

## FIG. 10

Translation of DNA ssCA216a

ArgArgArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAsp  
1 CCCGGCGTAGGTCGCGCAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGCTTCGCCG  
GGGCCGCATCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCACGCCGAAGCGGC

LeuMetGlyTyrIleProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAla  
61 ACCTCATGGGGTACATACCGCTCGTCGGCGCCCCTCTTGGAGGCGCTGCCAGGGCCCTGG  
TGGAGTACCCCATGTATGGCGAGCAGCCGCGGGGAGAACCTCCGCGACGGTCCCGGGACC

HisGlyValArgValLeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCys  
121 CGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAACATGCAACAGGGAACCTTCCTGGTT  
CGGTACCGCAGGCCCAAGACCTTCTGCCGCACTTGATACGTTGTCCCTTGAAGGACCAA

SerPheSerIlePheLeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyr  
181 GCTCTTCTCTATCTTCTTCTGGCCCTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCT  
CGAGAAAGAGATAGAAGGAAGACCGGGACGAGAGAACGAACAGTACACGGGCGAAGCCGGA

-----  
GlnValArgAsnSerThrGlyLeuTyrHisValThrAsnAspCysProAsnSerSerIle  
241 ACCAAGTGCGCAACTCCACGGGGCTTTACCACGTCACCAATGATTGCCCTAACTCGAGTA  
TGGTTCACGCGTTGAGGTGCCCCGAAATGGTGAGTGGTTACTAACGGGATTGAGCTCAT

-----overlap with CA167b-----  
ValTyrGluAlaAlaAspAlaIleLeuHisThrProGlyCysValProCysValArgGlu  
301 TTGTGTACGAAGCGGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCTGTG  
AACACATGCTTCGCCGGCTACGGTAGGACGTGTGAGGCCCCACGCAGGGAACGCAAGCAC

-----  
GlyAsnAlaSerArgCysTrpValAlaMetThrProThrValAla  
361 AGGGCAACGCCTCGAGGTGTTGGGTGGCGATGACCCCTACGGTGGCC  
TCCCGTTGCGGAGCTCCACAACCCACCGCTACTGGGGATGCCACCGG

SUBSTITUTE SHEET

## FIG. 11

Translation of DNA ssCA290a

LysLysAsnLysArgAsnThrAsnArgArgProGlnAspValLysPheProGlyGlyGly  
 1 AAAAAAAAAAACAACGTAACACCAACCGTCGCCCACAGGACGTCAAGTTCCCGGGTGGCG  
 TTTTTTTTTTGTTCATTGTGGTTGGCAGCGGGTGTCTGCAGTTCAAGGGCCACCGC

GlnIleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeuGlyValArgAla  
 61 GTCAGATCGTTGGTGGAGTTTACTTGTTCGCCGCGCAGGGGCCCTAGATTGGGTGTGCGCG  
 CAGTCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAACCCACACGCGC

ThrArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnProIleProLysAla  
 121 CGACGAGAAAGACTTCCGAGCGGTGCGAACCTCGAGGTAGACGCCAGCCTATCCCCAAGG  
 GCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCGGTGCGATAGGGGTTCC

ArgArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsn  
 181 CTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGTACCTTGGCCCCCTCTATGGCA  
 GAGCAGCCGGGCTCCCGTCTGGACCCGAGTCGGGCCCATGGGAACCGGGGAGATACCGT

GluGlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGly  
 241 ATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGGCCTAGCTGGG  
 TACTCCCGACGCCACCCGCCCTACCGAGGACAGAGGGGCACCGAGAGCCGGATCGACCC

-----  
 ProThrAspProArgArgArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCys  
 301 GCCCCACAGACCCCGGCGTAGGTGCGCGCAATTTGGGTAAGGTCATCGATACCCCTACGT  
 CGGGGTGTCTGGGGGCCGATCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCA

-----  
 GlyPheAlaAspLeuMetGlyTyrIleProLeuValGlyAlaProLeuGlyGlyAlaAla  
 361 GCGGCTTCGCGGACCTCATGGGGTACATACCGCTCGTCGGCGCCCCCTCTGGAGGCGCTG  
 CGCCGAAGCGGCTGGAGTACCCCATGTATGGCGAGCAGCCGCGGGGAGAACCTCCGCGAC

-----overlap with CA216a-----  
 ArgAlaLeuAlaHisGlyValArgValLeuGluAspGlyValAsnTyrAlaThrGlyAsn  
 421 CCAGGGCCCTGGCGCATGGCGTCCGGTTCTGGAAGACGGCGTGAACCTATGCAACAGGGA  
 GGTCCCGGGACCGGTACCGCAGGCCCAAGACCTTCTGCCGCACTTGATACGTTGTCCCT

-----  
 LeuProGlyCysSerPheSerThrPhe  
 481 ACCTTCCTGGTTGCTCTTTCTCTACCTTC  
 TGGAAGGACCAACGAGAAAGAGATGGAAG

SUBSTITUTE SHEET

11 / 40

Translation of DNA ag30a

## FIG. 12-1

#MetSerValValGlnProProGlyProProLeu

#MetAlaLeuValOP

- 1 CGCAGAAAGCGTCTAGCCATGGCGTTAGTATGAGTGTCTGTCAGCCTCCAGGACCCCCC  
GCGTCTTTTCGAGATCGGTACCGCAATCATACTACAGCACGTCTGGAGGTCTGGGGGGG  
ProGlyGluProAM
- 61 TCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGAC  
AGGGCCCTCTCGGTATCACCAGACGCCTTGGCCACTCATGTGGCCTTAACGGTCCTGCTG
- #MetProGlyAspLeuGlyValProProGlnAsp
- 121 CGGGTCCTTTCTTGGATCAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCCGCAAGA  
GCCCAGGAAAGAACCTAGTTGGGCGAGTTACGGACCTCTAAACCCGCACGGGGGCGTTCT  
CysAM OP AM GlyAlaCys  
\*
- 181 CTGCTAGCCGAGTAGTGTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTT  
GACGATCGGCTCATCACAAACCCAGCGCTTCCGGAACACCATGACGGACTATCCACGAA  
GluCysProGlyArgSerArgArgProCysThrMetSerThrAsnProLysProGlnLys
- 241 GCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAA  
CGCTCACGGGGCCCTCCAGAGCATCTGGCAGGTGGTACTCGTGCTTAGGATTTGGAGTTT  
LysAsnLysArgAsnThrAsnArgArgProGlnAspValLysPheProGlyGlyGlyGln
- 301 AAAAAACAAACGTAACACCAACCGTCGCCCACAGGACGTCAAGTTCCCGGTGGCGGTC  
TTTTTTTGTTCATTGTGGTTGGCAGCGGGTGTCTCTGCAGTTCAAGGGCCACCGCCAG  
IleValGlyGlyValTyrLeuLeuProArgArgGlyProArgLeuGlyValArgAlaThr
- 361 AGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGA  
TCTAGCAACCACCTCAAATGAACAACGGCGCGTCCCCGGGATCTAACCACACGCGCGCT  
ArgLysThrSerGluArgSerGlnProArgGlyArgArgGlnProIleProLysAlaArg
- 421 CGAGAAAGACTTCCGAGCGGTGCGAACCTCGAGGTAGACGTACAGCTATCCCCAAGGCTC  
GCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCATCTGCAGTTCGGATAGGGGTTCAG  
ArgProGluGlyArgThrTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGlu
- overlap with CA290a-----
- 481 GTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGCCCTCTATGGCAATG  
CAGCCGGGCTCCCGTCTGGACCCGAGTCGGGCCCATGGGAACGGGGAGATACCGTTAC  
GlyCysGlyTrpAlaGlyTrpLeuLeuSerProArgGlySerArgProSerTrpGlyPro
- 541 AGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGGCTCTCGGCCTAGCTGGGGCC  
TCCCGACGCCACCCGCCCTACCGAGGACAGAGGGGCACCGAGAGCCGGATCGACCCCGG  
ThrAspProArgArgArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCysGly

SUBSTITUTE SHEET

12 / 40

601 CCACAGACCCCGGCGTAGGTCGCGCAATTGGGTAAGGTCATCGATACCCTTACGTGCG  
GGTGTCTGGGGGCCGCATCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCACGC

Phe

-----  
661 GCTTC  
CGAAG

\* = Start of long HCV ORF  
| = Putative first amino acid of large HCV polyprotein  
# = Putative small encoded peptides (that may play a  
translational regulatory role)

FIG. 12-2

SUBSTITUTE SHEET

13/40

## FIG. 13

Translation of DNA CA205a

1 ValLeuGlyArgGluArgProCysGlyThrAlaOP AM GlyAlaCysGluCysProGly  
 GTCTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCAGAGTGGCCCGGG  
 CAGAACCCAGCGCTTTCGGAACACCATGACGGACTATCCACGAACGCTCACGGGGCCC

\*

61 ArgSerArgArgProCysThrMetSerThrAsnProLysProGlnArgLysThrLysArg  
 AGGTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGT  
 TCCAGAGCATCTGGCAGCTGGTACTCGTGCTTAGGATTTGGAGTTTCTTTTGGTTTGCA

121 AsnThrAsnArgArgProGlnAspValLysPheProGlyGlyGlyGlnIleValGlyGly  
 AACACCAACCGTCGCCCCACAGGACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGA  
 TTGTGGTTGGCAGCGGGTGTCTGCAGTTCAAGGGCCACCGCCAGTCTAGCAACCACCT

181 ValTyrLeuLeuProArgArgGlyProArgLeuGlyValArgAlaThrArgLysThrSer  
 GTTTACTTGTGCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCC  
 CAAATGAACAACGGCGCGTCCCCGGGATCTAACCACACGCGCGCTGCTCTTCTGAAGG

-----overlap with CA290a-----

241 GluArgSerGlnProArgGlyArgGlnProIleProLysAlaArgArgProGluGly  
 GAGCGGTGCGAACCTCGAGGTAGACGTCAGCCTATCCCCAAGGCTCGTCGGCCCGAGGGC  
 CTCGCCAGCGTTGGAGCTCCATCTGCAGTCGGATAGGGGTTCGAGCAGCCGGGCTCCCG

301 ArgThrTrpAlaGlnProGlyTyrProTrpProLeuTyrGlyAsnGluGlyCys  
 AGGACCTGGGCTCAGCCCGGTACCTTGGCCCCCTATGGCAATGAGGGCTGCG  
 TCCTGGACCCGAGTCGGGCCCATGGGAACCGGGAGATACCGTTACTCCCGACGC

\* = putative initiator methionine codon

SUBSTITUTE SHEET.



14 / 40

## FIG. 14

Translation of DNA 18g

```

#ProProOP
#SerThrMetAsnHisSerProValArgAsnTyrCysLeuHisAlaGluSerValAM Pro
#LeuHisHisGluSerLeuProCysGluGluLeuLeuSerSerArgArgLysArgLeuAla
1 CTCCACCATGAATCACTCCCCTGTGAGGAAGTACTGTCTTCACGCAGAAAGCGTCTAGCC
GAGGTGGTACTTAGTGAGGGGACACTCCTTGATGACAGAAGTGCCTCTTTTCGCAGATCGG

-----
#MetSerValValGlnProProGlyProProLeuProGlyGluProAM
MetAlaLeuValOP
61 ATGGCGTTAGTATGAGTGTCTGTCAGCCTCCAGGACCCCCCTCCCGGGAGAGCCATAGT
TACCGCAATCATACTCACAGCACGTCGGAGGTCCTGGGGGGGAGGGCCCTCTCGGTATCA

-----
121 GGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGATC
CCAGACGCCCTTGCCCACTCATGTGGCCTTAACGGTCCTGCTGGCCCAGGAAAGAACCCTAG

-----overlap with ag30a-----
#MetProGlyAspLeuGlyValProProGlnAspCysAM
181 AACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCGCAAGACTGCTAGCCGAGTAGTGT
TTGGGCGAGTTACGGACCTCTAAACCCGCACGGGGGCGTTCTGACGATCGGCTCATCACA

-----
OP AM GlyAlaCysGluCysProGlyArgSer
*
241 TGGGTCGCGAAAGGCCTTGTTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGT
ACCCAGCGCTTTCCGGAACACCATGACGGACTATCCACGAACGCTCACGGGGCCCTCCA

-----
ArgArg
301 CTCGTAGA
GAGCATCT

```

\* = Start of long HCV ORF

# = Putative small encoded peptides (that may play a translational regulatory role)

SUBSTITUTE SHEET

## FIG. 15

Translation of DNA 16jh

-----Overlap with 15e -----

1 GlyAlaCysTyrSerIleGluProLeuAspLeuProProIleIleGlnArgLeuHisGly  
 GGGGCCTGCTACTCCATAGAACCACTGGATCTACCTCCAATCATTCAAAGACTCCATGGC  
 CCCCGGACGATGAGGTATCTTGGTGACCTAGATGGAGGTTAGTAAGTTTCTGAGGTACCG

61 LeuSerAlaPheSerLeuHisSerTyrSerProGlyGluIleAsnArgValAlaAlaCys  
 CTCAGCGCATTTTCACTCCACAGTTACTCTCCAGGTGAAATTAATAGGGTGGCCGCATGC  
 GAGTCGCGTAAAAGTGAGGTGTCAATGAGAGGTCCACTTTAATTATCCCACCGGCGTACG

Gly\*  
G

121 LeuArgLysLeuGlyValProProLeuArgAlaTrpArgHisArgAlaArgSerValArg  
 CTCAGAAAACCTGGGGTACCGCCCTTGCGAGCTTGGAGACACCGGGCCCGGAGCGTCCGC  
 GAGTCTTTTGAACCCCATGGCGGGAACGCTCGAACCTCTGTGGCCCGGCGCTCGCAGGCG

181 AlaArgLeuLeuAlaArgGlyGlyArgAlaAlaIleCysGlyLysTyrLeuPheAsnTrp  
 GCTAGGCTTCTGGCCAGAGGAGGCAGGGCTGCCATATGTGGCAAGTACCTCTTCAACTGG  
 CGATCCGAAGACCGGTCTCCTCCGTCCCACGGTATACACCGTTCATGGAGAAGTTGACC

241 AlaValArgThrLysLeuLys  
 GCAGTAAGAACAAGCTCAAAC  
 CGTCATTCTTGTTCGAGTTTG

\* = nucleotide heterogeneity

SUBSTITUTE SHEET

16 / 40

COMBINED ORF OF DNAs p114a THROUGH 15e

FIG. 16-1

(p114a/CA167b/CA156e/CA84a/CA59a/K9-1/12f/14i/11b/7f/7e/  
8h/33c/40b/37b/35/36/81/32/33b/25c/14c/8f/33f/33g/39c/  
35f/19g/26g & 15e)

1 ArgSerArgAsnLeuGlyLysValIleAspThrLeuThrCysGlyPheAlaAspLeuMet  
AGGTCGCGCAATTTGGGTAAGGTACATCGATACCCTTACGTGCGGCTTCGCGGACCTCATG  
TCCAGCGCGTTAAACCCATTCCAGTAGCTATGGGAATGCACGCCGAAGCGGCTGGAGTAC

61 GlyTyrIleProLeuValGlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGly  
GGGTACATACCGCTCGTTCGGCGCCCTCTTGGAGGCGCTGCCAGGCGCTGGCGCATGGC  
CCCATGTATGGCGAGCAGCCGCGGGGAGAACCTCCGCGACGGTCCCGGGACCGGTACCG

121 ValArgValLeuGluAspGlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPhe  
GTCCGGGTTCTGGAAGACGCGTGAACATGCAACAGGGAACCTTCCTGGTTGCTCTTTC  
CAGGCCAAGACCTTCTGCCGCACTTGATACGTTGTCCCTTGAAGGACCAACGAGAAAG

181 SerIlePheLeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyrGlnVal  
TCTATCTTCTTCTGGCCCTGCTCTCTTGGTTGACTGTGCCCCGCTTCGGCCTACCAAGTG  
AGATAGAAGGAAGACCGGACGAGAGAACGAACGTGACACGGGCGAAGCCGGATGGTTCAC

241 ArgAsnSerThrGlyLeuTyrHisValThrAsnAspCysProAsnSerSerIleValTyr  
CGCAACTCCACGGGGCTTTACCACGTACCAATGATTGCCCTAACTCGAGTATGTGTAC  
GCGTTGAGGTGCCCCGAAATGGTGAGTGGTTACTAACGGGATTGAGTCTATAACACATG

301 GluAlaAlaAspAlaIleLeuHisThrProGlyCysValProCysValArgGluGlyAsn  
GAGGCGGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTGCTGAGGGCAAC  
CTCCGCGGGCTACGGTAGGACGTGTGAGGCCCCACGCAGGGAACGCAAGCACTCCCGTTG

361 AlaSerArgCysTrpValAlaMetThrProThrValAlaThrArgAspGlyLysLeuPro  
GCCTCGAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACCTCCCC  
CGGAGCTCCACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGG

421 AlaThrGlnLeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCysSerAla  
GCGACGCAGCTTCGACGTACATCGATCTGCTTGTGCGGAGCGCCACCTCTGTTTCGGCC  
CGCTGCGTCGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACAAGCCGG

481 LeuTyrValGlyAspLeuCysGlySerValPheLeuValGlyGlnLeuPheThrPheSer  
CTCTACGTGGGGGACCTATGCGGTCTGTCTTTCTTGTGCGGCCAACTGTTCACTTCTCT  
GAGATGCACCCCTGGATACGCCAGACAGAAAGAACAGCCGGTTGACAAGTGAAGAGA

541 ProArgArgHisTrpThrThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThr  
CCCAGGCGCCACTGGACGACGCAAGGTGCAATTGCTCTATCTATCCCGGCCATATAACG  
GGGTCCGCGGTGACCTGCTGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGC

601 GlyHisArgMetAlaTrpAspMetMetMetAsnTrpSerProThrThrAlaLeuValMet  
GGTCACCGCATGGCATGGGATATGATGATGAACCTGGTCCCTACGACGGCGTTGGTAATG  
CCAGTGGCGTACCGTACCTATACTACTTGTACAGGGGATGCTGCCGCAACCATTAC

661 AlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIleAlaGlyAlaHisTrpGly  
GCTCAGCTGCTCCGGATCCCAAGCCATCTTGGACATGATCGCTGGTGGTCACTGGGGA  
CGAGTCGACGAGGCCTAGGGTGTTCGGTAGAACCTGTACTAGCGACCACGAGTGACCCCT

721 ValLeuAlaGlyIleAlaTyrPheSerMetValGlyAsnTrpAlaLysValLeuValVal  
GTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACCTGGGCGAAGGTCTGGTAGTG  
CAGGACCGCCCGTATCGCATAAAGAGGTACCACCCCTTGACCGCTTCCAGGACCATCAC

781 LeuLeuLeuPheAlaGlyValAspAlaGluThrHisValThrGlyGlySerAlaGlyHis  
CTGCTGCTATTTGCCGGCGTGCAGCGGAAACCCACGTACCGGGGGAAGTGCCCGGCCAC  
GACGACGATAAACGGCCGAGCTGCGCCTTGGGTGCAGTGGCCCCCTTACGGCCGGTG

SUBSTITUTE SHEET

17 / 40

## FIG. 16-2

841 ThrValSerGlyPheValSerLeuLeuAlaProGlyAlaLysGlnAsnValGlnLeuIle  
ACTGTGTCTGGATTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAGCTGATC  
TGACACAGACCTAAACAATCGGAGGAGCGTGGTCCGCGGTTCTGCTTGCAGGTTCGACTAG

901 AsnThrAsnGlySerTrpHisLeuAsnSerThrAlaLeuAsnCysAsnAspSerLeuAsn  
AACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAACTGCAATGATAGCCTCAAC  
TTGTGGTTGCCGTCAACCGTGGAGTTATCGTGCCGGGACTTGACGTTACTATCGGAGTTG

961 ThrGlyTrpLeuAlaGlyLeuPheTyrHisHisLysPheAsnSerSerGlyCysProGlu  
ACCGGCTGGTTGGCAGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCTGAG  
TGGCCGACCAACCGTCCCGAAAAGATAGTGGTGTCAAGTTGAGAAGTCCGACAGGACTC

1021 ArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGlyProIleSerTyr  
AGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTAT  
TCCGATCGGTTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCCCGGGATAGTCAATA

1081 AlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrProProLysProCys  
GCCAACGGAAGCGGCCCCGACCAGCGCCCTACTGTGCGCACTACCCCCAAAACCTTGC  
CGTTGCCTTCGCCGGGGCTGGTCCGCGGGGATGACGACCGTGATGGGGGGTTTTGGAACG

1141 GlyIleValProAlaLysSerValCysGlyProValTyrCysPheThrProSerProVal  
GGTATTGTGCCCCGGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGCCCCGTG  
CCATAACACGGGGCGTTCTCACACACACCAGGCCATATAACGAAGTGAGGGTCGGGGCAC

1201 ValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGlyGluAsnAspThr  
GTGGTGGGAACGACCGACAGGTCCGGGCGCGCCACCTACAGCTGGGGTGAAAATGATACG  
CACCACCCTTGCTGGCTGTCCAGCCCGCGGGGTGGATGTGACCCCACTTTACTATGTC

1261 AspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPheGlyCysThrTrp  
GACGTCTTCGTCTTAACAATACCAGGCCACCGCTGGGCAATTGGTTTCGGTTGTACCTGG  
CTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCAAGCCAACATGGACC

1321 MetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysValIleGlyGlyAla  
ATGAACCTCAACTGGATTACCAAAGTGTGCGGAGCGCCTCCTTGTGTCTATCGGAGGGGCG  
TACTTGAGTTGACCTAAGTGGTTTCACACGCCCTCGCGGAGGAACACAGTAGCCTCCCCGC

1381 GlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisProAspAlaThrTyr  
GGCAACAACACCCTGCACTGCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATAC  
CCGTGTTGTGGGACGTGACGGGGTACTAACGAAGGCGTTTCGTAGGCTGCGGTGTATG

1441 SerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAspTyrProTyrArg  
TCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTTCGACTACCCGTATAGG  
AGAGCCACGCCGAGGCCAGGACCTAGTGTGGTCCACGGACAGCTGATGGGCATATCC

1501 LeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArgMetTyrValGly  
CTTTGGCATTATCCTTGTACCATCACTACACCATATTTAAATCAGGATGTACGTGGGA  
GAAACCGTAATAGGAACATGGTAGTTGATGTGGTATAAATTTAGTCTACATGCACCCCT

1561 GlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGluArgCysAspLeu  
GGGGTCGAACACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGCGAACGTTGCGATCTG  
CCCCAGCTTGTGTCGACCTTCGACGGACGTGACCTGCGCCCCGCTTGCAACGCTAGAC

1621 GluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThrGlnTrpGlnVal  
GAAGACAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACAGTGGCAGGTC  
CTTCTGTCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGATGTGTACCGTCCAG

1681 LeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIleHisLeuHisGln  
CTCCCGTGTTCCTTCAACAACCTACCAAGCCTTGTCCACCGGCCCTCATCCACCTCCACCAG  
GAGGGCACAAAGGAAGTGTGGGATGGTCCGAACAGGTGGCCGAGTAGGTGGAGGTGGTC

1741 AsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAlaSerTrpAlaIle  
AACATTGTGGACGTGCAGTACTGTACGGGTGGGGTCAAGCATCGCGTCTGGGCCATT  
TTGTAACACCTGCACGTATGAACATGCCCCACCCAGTTCGTAGCGCAGGACCCGGTAA

1801 LysTrpGluTyrValValLeuLeuPheLeuLeuAlaAspAlaArgValCysSerCys  
AAGTGGGAGTACGTGTTCTCCTGTTCTGCTTGACAGCGCGCGCTGCTCTCTGTC

SUBSTITUTE SHEET

18/40

## FIG. 16-3

TTCACCCCTCATGCAGCAAGAGGACAAGGAAGACGAACGTCTGCGCGCGCAGACGAGGACG  
1861 LeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeu  
TTGTGGATGATGCTACTCATATCCCAAGCGGAGGCGGCTTTGGAGAACCTCGTAATACTT  
AACACCTACTACGATGAGTATAGGGTTCCGCTCCGCCGAAACCTCTTGGAGCATTATGAA  
1921 AsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePheCysPhe  
AATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCTCGTGTCTTCTGCTTT  
TTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAA  
1981 AlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrp  
GCATGGTATTTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGGATGTGG  
CGTACCATAAACTTCCCATTCACCCACGGGCCTCGCCAGATGTGAAGATGCCCTACACC  
2041 ProLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluVal  
CCTCTCCTCCTGCTCCTGTTGGCGTTGCCCGAGCGGGCGTACGCGCTGGACACGGAGGTG  
GGAGAGGAGGACGAGGACAACCGCAACGGGGTCCGCCGATGCGCGACCTGTGCCTCCAC  
2101 AlaAlaSerCysGlyGlyValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyr  
GCCGCGTCGTGTGGCGGTGTTGTTCTCGTGGGTGATGGCGCTGACTCTGTCACCATAT  
CGGCGCAGCACACCGCCACAACAAGAGCAGCCCACTACCGCGACTGAGACAGTGGTATA  
2161 TyrLysArgTyrIleSerTrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGlu  
TACAAGCGCTATATCAGCTGGTGTGTTCTGTTGGTGGCTTCACTATTTTCTGACCAGAGTGAA  
ATGTTCCGATATAGTCGACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTT  
2221 AlaGlnLeuHisValTrpIleProProLeuAsnValArgGlyGlyArgAspAlaValIle  
GCGCAACTGCACGTGTGGATTCCCCCCTCAACGTCCGAGGGGGCGCGACGCCGTATC  
CGGTTGACGTGCACACCTAAGGGGGGGAGTTGCAGGCTCCCCCGCGCTGCGGCAGTAG  
2281 LeuLeuMetCysAlaValHisProThrLeuValPheAspIleThrLysLeuLeuLeuAla  
TTACTCATGTGTGCTGTACACCCGACTCTGGTATTGACATCACCAAAATGCTGCTGGCC  
AATGAGTACACACGACATGTGGGCTGAGACCATAAACTGTAGTGGTTTAAACGACGACCGG  
2341 ValPheGlyProLeuTrpIleLeuGlnAlaSerLeuLeuLysValProTyrPheValArg  
GTCTTCGGACCCCTTTGGATTCTTCAAGCCAGTTTGCTTAAAGTACCCTACTTTGTGCGC  
CAGAAGCCTGGGGAAACCTAAGAAGTTCGGTCAAACGAATTTCTATGGGATGAAACACGCG  
2401 ValGlnGlyLeuLeuArgPheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrVal  
GTCCAAGGCCCTTCTCCGTTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTG  
CAGGTTCCGGAAGAGGCCAAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCAC  
2461 GlnMetValIleIleLysLeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThr  
CAAATGGTCATCATTAAGTTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACT  
GTTTACCAGTAGTAATTCAATCCCCGGAATGACCGTGGATACAAATATTGGTAGAGTGA  
2521 ProLeuArgAspTrpAlaHisAsnGlyLeuArgAspLeuAlaValAlaValGluProVal  
CCTCTTCGGGACTGGGCGCACACGGCTTGCGAGATCTGGCCGTGGCTGTAGAGCCAGTC  
GGAGAAGCCCTGACCCGCGTGTGCGGAACGCTCTAGACCGGCACCGACATCTCGGTGAG  
2581 ValPheSerGlnMetGluThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGly  
GTCTTCTCCCAAATGGAGACCAAGCTCATCAGTGGGGGGCAGATACCGCCGCGTGGGT  
CAGAAGAGGGTTTACCTCTGGTTCGAGTAGTGCACCCCCGTCTATGGCGGCGCACGCCA  
2641 AspIleIleAsnGlyLeuProValSerAlaArgArgGlyArgGluIleLeuLeuGlyPro  
GACATCATCAACGGCTTGCCTGTTCCGCCCCGAGGGGCGGGAGATACTGCTCGGGCCA  
CTGTAGTAGTTGCCGAACGACAAAGCGGGCGTCCCCGGCCCTCTATGACGAGCCCGGT  
2701 AlaAspGlyMetValSerLysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGln  
CCGATGGAATGGTCTCCAAGGGGTGGAGGTGCTGGCGCCCATCACGGCGTACGCCAG  
CGGCTACCTTACCAGAGGTTCCCCACCTCCAACGACCGGGGTAGTGGCGCATGCGGGTC  
2761 GlnThrArgGlyLeuLeuGlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGln  
CAGACAAGGGGCTCCTAGGTTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAA  
GTCTGTTCCCCGGAGGATCCACGATTAGTGGTGGATTGACCGGCCCTGTTTTGGTT

SUBSTITUTE SHEET

19/40  
FIG. 16-4

2821 ValGluGlyGluValGlnIleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIle  
GTGGAGGGTGAGGTCCAGATTGTGTCAACTGCTGCCCAAACCTTCCTGGCAACGTGCATC  
CACCTCCCACTCCAGGTCTAACACAGTTGACGACGGGTTTGGAAGGACCGTTGCACGTAG

2881 AsnGlyValCysTrpThrValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLys  
AATGGGGTGTGCTGGACTGTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAG  
TTACCCACACGACCTGACAGATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTT

2941 GlyProValIleGlnMetTyrThrAsnValAspGlnAspLeuValGlyTrpProAlaPro  
GGTCCTGTCATCCAGATGTATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCCGTCCG  
CCAGGACAGTAGGTCTACATATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGC

3001 GlnGlySerArgSerLeuThrProCysThrCysGlySerSerAspLeuTyrLeuValThr  
CAAGGTAGCCGCTCATTTGACACCCTGCCTTGGGCTCCTCGGACCTTTACCTGGTTCAG  
GTTCCATCGGCGAGTAACGTGGGACGTGAACGCCGAGGAGCCTGGAAATGGACAGTGC

3061 ArgHisAlaAspValIleProValArgArgArgGlyAspSerArgGlySerLeuLeuSer  
AGGCACGCCGATGTTCATTCCTGCGCCGGCGGGGTGATAGCAGGGGCGAGCTGCTGTGCG  
TCCGTGCGGCTACAGTAAGGGCACGCGGCCGCCCACTATCGTCCCGCTCGGACGACAGC

3121 ProArgProIleSerTyrLeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGly  
CCCCGGCCCATTTCTACTTGAAAGGCTCCTCGGGGGTCCGCTGTTGTGCCCGCGGGG  
GGGGCCGGGTAAAGGATGAACCTTCCGAGGAGCCCCCAGGCGACAACACGGGGCGCCCC

3181 HisAlaValGlyIlePheArgAlaAlaValCysThrArgGlyValAlaLysAlaValAsp  
CACGCCGTGGGCATATTAGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGCGGGTGGAC  
GTGCGGCACCCGTATAAATCCCGCGCCACAGTGGGCACCTACCGATTCCGCCACCTG

3241 PheIleProValGluAsnLeuGluThrThrMetArgSerProValPheThrAspAsnSer  
TTTATCCCTGTGGAGAACCTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAAATCC  
AAATAGGGACACCTCTTGGATCTCTGTTGGTACTCCAGGGGCCACAAGTCCGATTAGG

3301 SerProProValValProGlnSerPheGlnValAlaHisLeuHisAlaProThrGlySer  
TCTCCACCAGTAGTGCCCGAGAGCTTCAGGTGGCTCACCTCCATGCCACAGGCAGC  
AGAGGTGGTCATCACGGGGTCTCGAAGGTCCACCGAGTGGAGGTACGAGGGTGTCCGTG

3361 GlyLysSerThrLysValProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeu  
GGCAAAGCACCAAGGTCCCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTC  
CCGTTTTCTGTGGTTCCAGGGCCGACGTATACGTCGAGTCCCGATATCCACGATCATGAG

3421 AsnProSerValAlaAlaThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIle  
AACCCCTCTGTGCTGCAACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATC  
TTGGGGAGACAACGACGTTGTGACCCGAAACACGAATGTACAGGTTCCGAGTACCCTAG

3481 AspProAsnIleArgThrGlyValArgThrIleThrThrGlySerProIleThrTyrSer  
GATCCTAACATCAGGACCGGGTGAGAACAAATTACCACTGGCAGCCCCATCACGTACTCC  
CTAGGATTGTAGTCCTGGCCCCACTCTTGTTAATGGTGACCGTGGGGTAGTGCATGAGG

3541 ThrTyrGlyLysPheLeuAlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIle  
ACCTACGGCAAGTTCTTGGCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATT  
TGGATGCCGTTCAAGGAACGGCTGCCGCCCCACGAGCCCCCGGAATACTGTATTATTAA

3601 CysAspGluCysHisSerThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAsp  
TGTGACGAGTGCCACTCCACGGATGCCACATCCATCTTGGGCATCGGCACTGTCTTGAC  
ACACTGCTCACGGTGAGGTGCTACGGTGTAGGTAGAACCCGTAGCCGTGACAGGAACCTG

3661 GlnAlaGluThrAlaGlyAlaArgLeuValValLeuAlaThrAlaThrProProGlySer  
CAAGCAGAGACTGCGGGGGCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCC  
GTTCTCTCTGACGCCCCGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGG

3721 ValThrValProHisProAsnIleGluGluValAlaLeuSerThrThrGlyGluIlePro  
GTCACGTGCCCCATCCCAACATCGAGGAGGTGCTCTGTCCACCACGGAGAGATCCCT  
CAGTGACACGGGGTAGGGTTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGA

3781 PheTyrGlyLysAlaIleProLeuGluValIleLysGlyGlyArgHisLeuIlePheCys  
TTTACGGCAAGGCTATCCCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTTCTGT

SUBSTITUTE SHEET

20/40

## FIG. 16-5

AAAATGCCGTTCCGATAGGGGAGCTTCATTAGTTCCTCCCCCTCTGTAGAGTAGAAGACA

3841 HisSerLysLysLysCysAspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAla  
CATTCAAAGAAGAAGTGCAGCAACTCGCCGCAAAGCTGGTCGCATTGGGCATCAATGCC  
GTAAGTTTCTTCTTCACGCTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGG

3901 ValAlaTyrTyrArgGlyLeuAspValSerValIleProThrSerGlyAspValValVal  
GTGGCCTACTACCGCGGTCTTGACGTGTCCGTCATCCCGACCAGCGGCGATGTTGTCTGTC  
CACCGGATGATGGCGCCAGAACTGCACAGGCAGTAGGGCTGGTCGCGCTACAACAGCAG

3961 ValAlaThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCys  
GTGGCAACCGATGCCCTCATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGC  
CACCGTTGGCTACGGGAGTACTGGCCGATATGGCCGCTGAAGCTGAGCCACTATCTGACC

4021 AsnThrCysValThrGlnThrValAspPheSerLeuAspProThrPheThrIleGluThr  
AATACGTGTGTACCCAGACAGTCGATTTCAGCCTTGACCCTACCTTCACCATTGAGACA  
TTATGCACACAGTGGGTCTGTGAGCTAAAGTCGGAAGTGGGATGGAAGTGGTAACCTCTGT

4081 IleThrLeuProGlnAspAlaValSerArgThrGlnArgArgGlyArgThrGlyArgGly  
ATCACGCTCCCCAGGATGCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGG  
TAGTGCAGAGGGGTCTACGACAGAGGGCGTGAGTTGCAGCCCCGTCTGACCGTCCCC

4141 LysProGlyIleTyrArgPheValAlaProGlyGluArgProSerGlyMetPheAspSer  
AAGCCAGGCATCTACAGATTGTGGCAGCGGGGAGCGCCCTCCGGCATGTTCCACTCG  
TTCGGTCCGTAGATGTCTAAACACCGTGGCCCCCTCGCGGGAGGCGGTACAAGCTGAGC

4201 SerValLeuCysGluCysTyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGlu  
TCCGTCTCTGTGAGTGCTATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCCGCCGAG  
AGGCAGGAGACACTCACGATACTGCGTCCGACACGAACCATACTCGAGTGCGGGCGGCTC

4261 ThrThrValArgLeuArgAlaTyrMetAsnThrProGlyLeuProValCysGlnAspHis  
ACTACAGTTAGGCTACGAGCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCAT  
TGATGTCAATCCGATGCTCGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCTCTGGTA

4321 LeuGluPheTrpGluGlyValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSer  
CTTGAATTTTGGGAGGGCGTCTTACAGGCCTCACTCATATAGATGCCACTTTCTATCC  
GAACTTAAACCCCTCCCGCAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGG

4381 GlnThrLysGlnSerGlyGluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCys  
CAGACAAAGCAGAGTGGGGAGAACCTTCTTACCTAGGTAGCGTACCAAGCCACCGTGTGC  
GTCTGTTTCGTCTACCCCTCTTGAAGGAATGGACCATCGCATGGTTCCGTGGCACACG

4441 AlaArgAlaGlnAlaProProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeu  
GCTAGGGCTCAAGCCCCCTCCCCATCGTGGGACCAGATGTGGAAGTGTGTTGATTGCGCTC  
CGATCCCGAGTTCGGGGAGGGGTAGCACCTGGTCTACACCTTCAAACTAAGCGGAG

4501 LysProThrLeuHisGlyProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGlu  
AAGCCCACCTCCATGGGCCAACCCCTGCTATACAGACTGGGCGCTGTTTCAAGATGAA  
TTCGGGTGGGAGGTACCGGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTT

4561 IleThrLeuThrHisProValThrLysTyrIleMetThrCysMetSerAlaAspLeuGlu  
ATCACCTGACGCACCCAGTCACAAATACATCATGACATGCATGTCGGCCGACCTGGAG  
TAGTGGGACTGCGTGGGTCACTGGTTTATGTAGTACTGTACGTACAGCCGCTGGACCTC

4621 ValValThrSerThrTrpValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCys  
GTCGTACAGAGCACCTGGGTGCTCGTTGGCGGCTCCTGGCTGCTTTGGCCGCGTATTGC  
CAGCAGTGCTCGTGACCCACGAGCAACCGCCGAGGACCGACGAAACCGCGCATAACG

4681 LeuSerThrGlyCysValValIleValGlyArgValValLeuSerGlyLysProAlaIle  
CTGTCAACAGGCTGCGTGGTCATAGTGGGCAGGGTCTGCTTGTCCGGGAAGCCGGCAATC  
GACAGTTGTCCGACGCACCACTATCACCCGTCCAGCAGAACAGGCCCTTCGGCCGTTAG

4741 IleProAspArgGluValLeuTyrArgGluPheAspGluMetGluGluCysSerGlnHis  
ATACCTGACAGGGAAGTCTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCAC  
TATGGACTGTCCCTTACAGAGATGGCTCTCAAGCTACTCTACCTTCTCAGAGAGTCGTG

SUBSTITUTE SHEET

21 / 40

## FIG. 16-6

4801 LeuProTyrIleGluGlnGlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGly  
TTACCGTACATCGAGCAAGGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGC  
AATGGCATGTAGCTCGTTCCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCG

4861 LeuLeuGlnThrAlaSerArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrp  
CTCCTGCAGACCGCTCCCGTCAGGCAGAGGTTATCGCCCCCTGCTGTCCAGACCAACTGG  
GAGGACGTCTGGCGCAGGGCAGTCCGTCTCCAATAGCGGGACGACAGGTCTGGTTGACC

4921 GlnLysLeuGluThrPheTrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyr  
CAAAACTCGAGACCTTCTGGGCGAAGCATATGTGGAACCTCATCAGTGGGATACAATAC  
GTTTTTGAGCTCTGGAAGACCCGCTTCGTATACACCTTGAAGTAGTCACCCTATGTTATG

4981 LeuAlaGlyLeuSerThrLeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThr  
TTGGCGGGCTTGTCACGCTGCCTGGTAACCCGCCATTGCTTCATTGATGGCTTTTACA  
AACCGCCCGAACAGTTGCGACGGACCATTGGGGCGGTAACGAAGTAACTACCGAAAATGT

5041 AlaAlaValThrSerProLeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGly  
GCTGCTGTACCAGCCCACTAACCCTAGCCAAACCCTCCTCTTCAACATATTGGGGGGG  
CGACGACAGTGGTCGGGTGATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCC

5101 TrpValAlaAlaGlnLeuAlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeu  
TGGGTGGCTGCCAGCTCGCCGCCCCCGGTGCCGCTACTGCCCTTGTGGGCGCTGGCTTA  
ACCCACCGACGGTTCGAGCGGGCGGGGCCACGGCGATGACGGAACACCCGCGACCGAAT

5161 AlaGlyAlaAlaIleGlySerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGly  
GCTGGCGCCGCCATCGGCAGTGTGGACTGGGGAAGGTCTCTATAGACATCCTTGAGGG  
CGACCGCGCGGTAGCCGTCACAACCTGACCCCTTCAGGAGTATCTGTAGGAACGTCC

5221 TyrGlyAlaGlyValAlaGlyAlaLeuValAlaPheLysIleMetSerGlyGluValPro  
TATGGCGCGGGCGTGGCGGGAGCTCTTGTGGCATTCAAGATCATGAGCGGTGAGGTCCCC  
ATACCGCGCCCGCACCGCCCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGG

5281 SerThrGluAspLeuValAsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValVal  
TCCACGGAGGACCTGGTCAATCTACTGCCCGCCATCCTCTCGCCCGGAGCCCTCGTAGTC  
AGGTGCCTCCTGGACAGTTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAG

5341 GlyValValCysAlaAlaIleLeuArgArgHisValGlyProGlyGluGlyAlaValGln  
GGCGTGGTCTGTGCAGCAATACTGCGCCGGCACGTTGGCCCGGGCGAGGGGGCAGTGCAG  
CCGCACCAGACACGTCGTTATGACGCGGCCGTGCAACCGGGCCCGCTCCCCCGTCACGTC

5401 TrpMetAsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValSerProThrHisTyr  
TGGATGAACCGGCTGATAGCCTTCGCCTCCCGGGGAACCATGTTTCCCCCAGCACTAC  
ACCTACTTGGCCGACTATCGGAAGCGGAGGGCCCCCTTGGTACAAAGGGGGTGCGTGATG

5461 ValProGluSerAspAlaAlaAlaArgValThrAlaIleLeuSerSerLeuThrValThr  
GTGCCGGAGAGCGATGCAGCTGCCCCGCTCACTGCCATACTCAGCAGCCTCACTGTAACC  
CACGGCCTCTCGCTACGTCGACGGGCGCAGTGACGGTATGAGTCGTGAGTGACATTGG

5521 GlnLeuLeuArgArgLeuHisGlnTrpIleSerSerGluCysThrThrProCysSerGly  
CAGCTCCTGAGGCGACTGCACAGTGGATAAGCTCGGAGTGTAACCTCCATGCTCCGGT  
GTCGAGGACTCCGCTGACGTGGTCACCTATTGAGCCTCACATGGTGAGGTACGAGGCCA

5581 SerTrpLeuArgAspIleTrpAspTrpIleCysGluValLeuSerAspPheLysThrTrp  
TCCTGGCTAAGGGACATCTGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGG  
AGGACCGATTCCCTGTAGACCCTGACCTATACGCTCCACAACCTCGTGAAATTCTGGACC

5641 LeuLysAlaLysLeuMetProGlnLeuProGlyIleProPheValSerCysGlnArgGly  
CTAAAGCTAAGCTCATGCCACAGCTGCCTGGGATCCCCCTTGTGTCTGCCAGCGCGGG  
GATTTTCGATTTCGAGTACGGTGTGACGGACCCTAGGGGAAACACAGGACGGTTCGCGCCC

5701 TyrLysGlyValTrpArgValAspGlyIleMetHisThrArgCysHisCysGlyAlaGlu  
TATAAGGGGGTCTGGCGAGTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAG  
ATATTCCCCCAGACCGCTCACCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTC

5761 IleThrGlyHisValLysAsnGlyThrMetArgIleValGlyProArgThrCysArgAsn  
ATCACTGGACATGTCAAAACGGGACGATGAGGATCGTCCGTCTAGGACCTGCAGGAAC

SUBSTITUTE SHEET



22/40

## FIG. 16-7

TAGTGACCTGTACAGTTTTTGGCCTGCTACTCCTAGCAGCCAGGATCCTGGACGTCCTTG

5821 MetTrpSerGlyThrPheProIleAsnAlaTyrThrThrGlyProCysThrProLeuPro  
ATGTGGAGTGGGACCTTCCCCATTAAATGCCACACACGGGCCCCGTGTACCCCCCTTCCCT  
TACACCTCACCTGGAAGGGTAATTACGGATGTGGTGCCCGGGACATGGGGGAAGGA

5881 AlaProAsnTyrThrPheAlaLeuTrpArgValSerAlaGluGluTyrValGluIleArg  
GCGCCGAACCTACAGTTCGCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGG  
CGCGGCTTGATGTGCAAGCGGATACCTCCACAGACGTCTCCTTATACACCTCTATTCC

5941 GlnValGlyAspPheHisTyrValThrGlyMetThrThrAspAsnLeuLysCysProCys  
CAGGTGGGGGACTTCCACTACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGC  
GTCCACCCCTGAAGGTGATGCACTGCCATCTGATGACTGTTAGAGTTTACGGGCACG

6001 GlnValProSerProGluPhePheThrGluLeuAspGlyValArgLeuHisArgPheAla  
CAGGTCCCATCGCCGAATTTTTTACAGAATTGGACGGGGTGCCTTACATAGGTTTGGCG  
GTCCAGGGTAGCGGGCTTAAAAAGTGCTTAACCTGCCCCACGGGATGTATCCAAACGC

6061 ProProCysLysProLeuLeuArgGluGluValSerPheArgValGlyLeuHisGluTyr  
CCCCCTGCAAGCCCTTGCTGCGGGAGGAGGTATCATTAGAGTAGGACTCCACGAATAC  
GGGGGACGTTTCGGGAACGACGCCCTCTCCATAGTAAGTCTCATCTGAGGTGCTTATG

6121 ProValGlySerGlnLeuProCysGluProGluProAspValAlaValLeuThrSerMet  
CCGGTAGGGTCGCAATTACCTTTCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATG  
GGCCATCCCAGCGTTAATGGAACGCTCGGGCTTGGCCTGCACCGGCACAACGTCAGGTAC

6181 LeuThrAspProSerHisIleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySer  
CTCACTGATCCCTCCCATATAACAGCAGAGGCGGCGGGCGAAGGTTGGCGAGGGGATCA  
GAGTGACTAGGGAGGGTATATTGTCGTCTCCGCGGCGCGCTTCAAACCGCTCCCCTAGT

6241 ProProSerValAlaSerSerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThr  
CCCCCTCTGTGGCCAGCTCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACT  
GGGGGAGACACCGGTTCGAGGAGCCGATCGGTTCGATAGGCGAGGTAGAGAGTTCCGTTGA

6301 CysThrAlaAsnHisAspSerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArg  
TGCACCGCTAACCATGACTCCCTGATGCTGAGCTCATAGAGGCCAACCTCCTATGGAGG  
ACGTGGCGATTGGTACTGAGGGGACTACGACTCGAGTATCTCCGTTGGAGGATACCTCC

6361 GlnGluMetGlyGlyAsnIleThrArgValGluSerGluAsnLysValValIleLeuAsp  
CAGGAGATGGGCGGCAACATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGAC  
GTCCTCTACCCGCGTTGTAGTGGTCCCAACTCAGTCTTTTGTTCACCACTAAGACCTG

6421 SerPheAspProLeuValAlaGluGluAspGluArgGluIleSerValProAlaGluIle  
TCCTTCGATCCGCTTGTGGCGGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATC  
AGGAAGCTAGGCGAACACCGCCTCCTCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAG

6481 LeuArgLysSerArgArgPheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsn  
CTGCGGAAGTCTCGGAGATTGCGCCAGGCCCTGCCGTTTGGGCGCGGCCGACTATAAC  
GACGCCCTCAGAGCCTCTAAGCGGGTCCGGGACGGGCAAACCCGCGCGCCTGATATTG

6541 ProProLeuValGluThrTrpLysLysProAspTyrGluProProValValHisGlyCys  
CCCCGCTAGTGGAGACGTGGAAGAGCCGACTACGAACCACTGTGGTCCATGGCTGT  
GGGGGCGATCACCTCTGCACCTTTTTCGGGCTGATGCTTGGTGGACACCAGGTACCGACA

6601 ProLeuProProProLysSerProProValProProProArgLysLysArgThrValVal  
CCGCTTCCACCTCAAAGTCCCTCTGTGCCTCCGCTCGGAAGAAGCGGACGGTGGTC  
GGCGAAGGTGGAGGTTTCAGGGGAGGACACGGAGGCGGAGCCTTCTTCGCTGCCACCAAG

6661 LeuThrGluSerThrLeuSerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySer  
CTCACTGAATCAACCCTATCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGC  
GAGTGACTTAGTTGGGATAGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCG

6721 SerSerThrSerGlyIleThrGlyAspAsnThrThrThrSerSerGluProAlaProSer  
TCCTCAACTTCCGGCATTACGGGCGACAATACGACAACATCTCTGAGCCCGCCCTTCT  
AGGAGTTGAAGGCGTAATGCCGCTGTTATGCTGTTGTAGGAGACTCGGGCGGGGAAGA

SUBSTITUTE SHEET

23/40

## FIG. 16-8

6781 GlyCysProProAspSerAspAlaGluSerTyrSerSerMetProProLeuGluGlyGlu  
GGCTGCCCCCGACTCCGACGCTGAGTCTTATCTCTCCATGCCCCCTGGAGGGGGAG  
CCGACGGGGGGGCTGAGGCTGCGACTCAGGATAAGGAGGTACGGGGGGGACCTCCCCCTC

6841 ProGlyAspProAspLeuSerAspGlySerTrpSerThrValSerSerGluAlaAsnAla  
CCTGGGGATCCGGATCTTAGCGACGGGTATGGTCAACGGTCAGTAGTGAGGCCAACGCG  
GGACCCCTAGGCCTAGAATCGCTGCCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGC

6901 GluAspValValCysCysSerMetSerTyrSerTrpThrGlyAlaLeuValThrProCys  
GAGGATGTCTGTGTCTCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCGTGC  
CTCTACAGCACACGACGAGTTACAGAATGAGAACCTGTCCGCGTGAGCAGTGGGGCAGC

6961 AlaAlaGluGluGlnLysLeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHis  
GCCGCGGAAGAACAGAACTGCCCCATCAATGCCTAAGCAACTCGTTGCTACGTACCCAC  
CGGCGCCTTCTTGTCTTTGACGGGTAGTTACGTGATTCTGTTAGCAACGATGCAGTGGT

7021 AsnLeuValTyrSerThrThrSerArgSerAlaCysGlnArgGlnLysLysValThrPhe  
AATTGGGTGATTCCACCACCTCAGCAGTGCCTTGCCAAAGGCAGAGAAAGTCACATTT  
TTAAACCACATAAGGTGGTGGAGTGCGTCACGAACGGTTTCCGTCTTCTTCAGTGTA

7081 AspArgLeuGlnValLeuAspSerHisTyrGlnAspValLeuLysGluValLysAlaAla  
GACAGACTGCAAGTTCTGGACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCG  
CTGTCTGACGTTCAAGACCTGTGCGTAATGGTCCTGCATGAGTTCTCCAATTTCTGTCGC

7141 AlaSerLysValLysAlaAsnLeuLeuSerValGluGluAlaCysSerLeuThrProPro  
GCGTCAAAAGTGAAGGCTAACTTGCTATCCGTAGAGGAAGCTTGACGCTGACGCCCCCA  
CGCAGTTTCACTTCCGATTGAACGATAGGCATCTCTTCGAACGTGGACTGCGGGGGT

7201 HisSerAlaLysSerLysPheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLys  
CACTCAGCCAAATCCAAGTTTGGTTATGGGGCAAAGACGTCCGTTGCCATGCCAGAAAG  
GTGAGTCGGTTTAGGTTCAAACCAATACCCCGTTTCTGTCAGGCAACGGTACGGTCTTTC

7261 AlaValThrHisIleAsnSerValTrpLysAspLeuLeuGluAspAsnValThrProIle  
GCCGTAACCCACATCAACTCCGTGTGGAAAGACCTTCTGGAAGACAATGTAACACCAATA  
CGGCATTGGGTGTAGTTGAGGCACACCTTCTGGAAGACCTTCTGTTACATTGTGGTTAT

7321 AspThrThrIleMetAlaLysAsnGluValPheCysValGlnProGluLysGlyGlyArg  
GACACTACCATCATGGCTAAGAACGAGGTTTTCTGCGTTCAGCCTGAGAAGGGGGGTCTGT  
CTGTGATGGTAGTACCGATTCTTGCTCCAAAGACGCAAGTCGGACTCTTCCCCCAGCA

7381 LysProAlaArgLeuIleValPheProAspLeuGlyValArgValCysGluLysMetAla  
AAGCCAGCTCGTCTCATCGTTTCCCGATCTGGGCGTGCAGCGTGTGCGAAAAGATGGCT  
TTCGGTCGAGCAGAGTAGCACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGA

7441 LeuTyrAspValValThrLysLeuProLeuAlaValMetGlySerSerTyrGlyPheGln  
TTGTACGACGTGGTTACAAAGCTCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAA  
AACATGCTGCACCAATGTTTCGAGGGGAACCGGCACTACCTTCGAGGATGCCTAAGGTT

7501 TyrSerProGlyGlnArgValGluPheLeuValGlnAlaTrpLysSerLysLysThrPro  
TACTCACCAGGACAGCGGTTGAATCTCTCGTCAAGCGTGAAGTCCAAGAAAACCCCA  
ATGAGTGGTCTGTGCCCCAATTAAGGAGCACGTTTCGCACCTTCAGGTTCTTTTGGGGT

7561 MetGlyPheSerTyrAspThrArgCysPheAspSerThrValThrGluSerAspIleArg  
ATGGGGTCTCGTATGATACCCGTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGT  
TACCCCAAGAGCATACTATGGGCGACGAACTGAGGTGTCAGTACTCTCGCTGTAGGCA

7621 ThrGluGluAlaIleTyrGlnCysCysAspLeuAspProGlnAlaArgValAlaIleLys  
ACGGAGGAGGCAATCTACCAATGTTGTGACCTCGACCCCAAGCCCGTGGCCATCAAG  
TGCTCTCCGTTAGATGGTTACAACACTGGAGCTGGGGGTTGGGGCGACCGGTAGTTC

7681 SerLeuThrGluArgLeuTyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCys  
TCCCTCACCAGAGAGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGC  
AGGGAGTGGCTCTCCGAAATACAACCCCGGGAGAATGGTTAAGTTCCCCCTCTTGACG

7741 GlyTyrArgArgCysArgAlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThr  
GGCTATCGCAGGTGCCGCGAGCGGCTACTGACAACCTAGCTGTGGTAACACCCTCACT

SUBSTITUTE SHEET

## FIG. 16-9

CCGATAGCGTCCACGGCGCGCTCGCCGCATGACTGTTGATCGACACCATTTGTGGGAGTGA

7801 CysTyrIleLysAlaArgAlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeu  
TGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTCCAGGACTGCACCATGCTC  
ACGATGTAGTTCCGGGCCCCGTCGGACAGCTCGGCGTCCCAGGTCCTGACGTGGTACGAG

7861 ValCysGlyAspAspLeuValValIleCysGluSerAlaGlyValGlnGluAspAlaAla  
GTGTGTGGCGACGACTTAGTCGTTATCTGTGAAAGCGCGGGGTCCAGGAGGACGCGGCG  
CACACCCGCTGCTGAATCAGCAATAGACACTTTCGCGCCCCCAGGTCTCTCGCGCCG

7921 SerLeuArgAlaPheThrGluAlaMetThrArgTyrSerAlaProProGlyAspProPro  
AGCCTGAGAGCCTTCACGGAGGCTATGACCAGGTACTCCGCCCCCTGGGGACCCCCA  
TCGGACTCTCGGAAGTGCTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGGT

7981 GlnProGluTyrAspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaHis  
CAACCAGAATACGACTTGGAGCTCATAACATCATGCTCTCCAACGTGTGAGTCGCCCCAC  
GTTGGTCTTATGCTGAACCTCGAGTATTGTAGTACGAGGAGGTGCACAGTCAGCGGGT

8041 AspGlyAlaGlyLysArgValTyrTyrLeuThrArgAspProThrThrProLeuAlaArg  
GACGGCGCTGGAAAGAGGGTCTACTACCTCACCCGTGACCCCTACAACCCCCCTCGCGAGA  
CTGCCGCGACCTTTCTCCAGATGATGGAGTGGGCACTGGGATGTTGGGGGAGCGCTCT

8101 AlaAlaTrpGluThrAlaArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMet  
GCTGCGTGGGAGACAGCAAGACACACTCCAGTCAATTCCTGGCTAGGCAACATAATCATG  
CGACGCACCCTCTGTCGTTCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTAC

8161 PheAlaProThrLeuTrpAlaArgMetIleLeuMetThrHisPhePheSerValLeuIle  
TTTGCCCCACACTGTGGGCGAGGATGATACTGATGACCCATTCTTTAGCGTCCCTTATA  
AAACGGGGGTGTGACACCCGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATAT

8221 AlaArgAspGlnLeuGluGlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIle  
GCCAGGGACCAGCTTGAACAGGCCCTCGATTGCGAGATCTACGGGGCCTGCTACTCCATA  
CGGTCCCTGGTCGAACCTGTCCGGGAGCTAACGCTCTAGATGCCCCGACGATGAGGTAT

8281 GluProLeuAspLeuProProIleIleGlnArgLeu  
GAACCACTTGATCTACCTCCAATCATTCAAAGACTC  
CTTGGTGAAC TAGATGGAGGTTAGTAAGTTTCTGAG

SUBSTITUTE SHEET

25 / 40

## FIG. 17-1

-319 CACTCCACCATGAATCACTCCCCTGTGAGGAACACTGTCTTCACGCAGAAAGCGTCTAG  
GTGAGGTGGTACTTAGTGAGGGGACACTCCTTGATGACAGAAGTGCCTCTTTCGCAGATC

-259 CCATGGCGTTAGTATGAGTGTCTGTCAGCCCTCCAGGACCCCCCTCCCGGGAGAGCCATA  
GGTACCGCAATCATACTACAGCACGTCTGGAGGTCTTGGGGGGAGGGCCCTCTCGGTAT

-199 GTGGTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCCCTTTCTTGGA  
CACCAGACGCCCTTGGCCACTCATGTGGCCTTAACGGTCTGCTGGCCAGGAAAGAACCT

-139 TCAACCCGCTCAATGCCTGGAGATTTGGGCGTGCCCCGCAAGACTGCTAGCCGAGTAGT  
AGTTGGGCGAGTTACGGACCTCTAAACCCGCACGGGGCGTTCTGACGATCGGCTCATCA

- 79 GTTGGGTCTGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAG  
CAACCCAGCGCTTTCGGAACACCATGACGGACTATCCACGAACGCTCACGGGGCCCTC

- 19 GTCTCGTAGACCGTGCACC  
CAGAGCATCTGGCACGTGG

Arg Thr  
1 MetSerThrAsnProLysProGlnLysLysAsnLysArgAsnThrAsnArgArgProGln  
ATGAGCACGAATCCTAAACCTCAAAAAAACAACCGTAACACCAACCGTCCGCCACAG  
TACTCGTGCTTAGGATTTGGAGTTTTTTTTTTGTTTGCATTGTGGTTGGCAGCGGGTGTCT

61 AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValTyrLeuLeuProArgArg  
GACGTCAAGTTCCCGGGTGGCGGTCTAGTCTGGTGGAGTTTACTTGTGCGCGCAGG  
CTGCAGTTCAAGGGCCACCGCCAGTCTAGCAACCACCTCAATGAACAACGGCGCGTCC

121 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly  
GGCCCTAGATTGGGTGTGCGCGCGACGAGAAAGACTTCCGAGCGGTCTGCAACCTCGAGGT  
CCGGGATCTAACCACACGCGCGCTGCTCTTTCTGAAGGCTCGCCAGCGTTGGAGCTCCA

181 ArgArgGlnProIleProLysAlaArgArgProGluGlyArgThrTrpAlaGlnProGly  
AGACGTCAGCCTATCCCAAGGCTCGTCTGGCCCCGAGGGCAGGACCTGGGCTCAGCCCGGG  
TCTGCAGTCGGATAGGGGTTCGAGCAGCCGGGCTCCCGTCTGGACCCGAGTCTGGGGCC

241 TyrProTrpProLeuTyrGlyAsnGluGlyCysGlyTrpAlaGlyTrpLeuLeuSerPro  
TACCCTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC  
ATGGGAACCGGGGAGATACCGTTACTCCCGACGCCACCCGCCCTACCGAGGACAGAGGG

301 ArgGlySerArgProSerTrpGlyProThrAspProArgArgArgSerArgAsnLeuGly  
CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCGCGGTAGGTCTGCGCAATTTGGGT  
GCACCGAGAGCCGGATCGACCCCGGGGTGTCTGGGGGCCGATCCAGCGCGTTAAACCCA

361 LysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProLeuVal  
AAGGTCATCGATAACCCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTC  
TTCCAGTAGCTATGGGAATGCACGCCGAAGCGGTGGAGTACCCCATGTATGGCGAGCAG

421 GlyAlaProLeuGlyGlyAlaAlaArgAlaLeuAlaHisGlyValArgValLeuGluAsp  
GGCGCCCCCTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGAC  
CCGCGGGGAGAACCTCCGCGACGGTCCCGGGACCGGTACCGCAGGCCCAAGACCTTCTG

Thr  
481 GlyValAsnTyrAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeuAla  
GGCGTGAACCTATGCAACAGGGAACCTTCCTGGTTGCTCTTCTCTATCTTCTTCTGGCC  
CCGCACTTGATACGTTGTCCCTTGAAGGACCAACGAGAAAGAGATAGAAGGAAGACCGG

541 LeuLeuSerCysLeuThrValProAlaSerAlaTyrGlnValArgAsnSerThrGlyLeu  
CTGCTCTCTTGCTTGACTGTGCCCCGCTTCGGCCTACCAAGTGCAGCAACTCCACGGGGCTT  
GACGAGAGAACGAACGTACACGGGCGAAGCCGGATGGTTACGCGTTGAGGTGCCCGGAA

601 TyrHisValThrAsnAspCysProAsnSerSerIleValTyrGluAlaAlaAspAlaIle  
TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGCGATGCCATC  
ATGGTGCAGTGGTTACTAACGGGATTGAGCTCATAACACATGCTCCGCCGGCTACGGTAG

661 LeuHisThrProGlyCysValProCysValArgGluGlyAsnAlaSerArgCysTrpVal  
CTGCACACTCCGGGGTGGTCCCTTGGCTTCGTGAGGGCAACGCCTCGAGGTGTTGGGTG  
GACGTGTGAGGCCCCACGCAGGGAACGCAAGCACTCCCGTTGCGGAGCTCCACAACCCAC

SUBSTITUTE SHEET

26/40

## FIG. 17-2

- 721 AlaMetThrProThrValAlaThrArgAspGlyLysLeuProAlaThrGlnLeuArgArg  
GCGATGACCCCTACGGTGGCCACCAGGGATGGCAAACCTCCCGCGACGCAGCTTCGACGT  
CGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTTGAGGGGGCTGCGTCAAGCTGCA
- 781 HisIleAspLeuLeuValGlySerAlaThrLeuCysSerAlaLeuTyrValGlyAspLeu  
CACATCGATCTGCTTGTCTGGGAGCGCCACCCTCTGTTCCGGCCCTCTACGTGGGGGACCTA  
GTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACAAGCCGGGAGATGCACCCCTGGAT
- 841 CysGlySerValPheLeuValGlyGlnLeuPheThrPheSerProArgArgHisTrpThr  
TGCGGGTCTGTCTTTCTTGTCTGGCCAACTGTTACCTTCTCTCCAGGCGCCACTGGACG  
ACGCCCAGACAGAAAGAACAGCCGGTTGACAAGTGAAGAGAGGGTCCGCGGTGACCTGC
- 901 ThrGlnGlyCysAsnCysSerIleTyrProGlyHisIleThrGlyHisArgMetAlaTrp  
ACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCAACGCATGGCATGG  
TGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTATATTGCCAGTGGCGTACCGTACC
- 961 Val  
AspMetMetMetAsnTrpSerProThrThrAlaLeuValMetAlaGlnLeuLeuArgIle  
GATATGATGATGAACCTGGTCCCTACGACGGCGTTGGTAATGGCTCAGCTGCTCCGGATC  
CTATACTACTACTTGACCAGGGGATGCTGCCGCAACCATACCGAGTCGACGAGGCCCTAG
- 1021 ProGlnAlaIleLeuAspMetIleAlaGlyAlaHisTrpGlyValLeuAlaGlyIleAla  
CCACAAGCCATCTTGGACATGATCGCTGGTGTCTACTGGGGAGTCTTGGCGGGCATAGCG  
GGTGTTCGGTAGAACCTGTACTAGCGACCAAGTGAACCCCTCAGGACCGCCCGTATCGC
- 1081 TyrPheSerMetValGlyAsnTrpAlaLysValLeuValValLeuLeuPheAlaGly  
TATTTCTCCATGGTGGGGAACCTGGGCGAAGGTCTGTAGTGTCTGTCTATTTGCCGGC  
ATAAAGAGGTACCACCCCTTGACCCGCTTCCAGGACCATCAGACGACGATAAACGGCCG
- 1141 ValAspAlaGluThrHisValThrGlyGlySerAlaGlyHisThrValSerGlyPheVal  
GTCGACGCGGAAACCCACGTCAACGGGGGAAGTGCCGGCCACACTGTGTCTGGATTTGTT  
CAGCTGCGCCTTTGGGTGCAGTGGCCCCCTTACGGCCGGTGTGACACAGACCTAAACAA
- 1201 SerLeuLeuAlaProGlyAlaLysGlnAsnValGlnLeuIleAsnThrAsnGlySerTrp  
AGCCTCTCTCGACCAAGGCGCCAGCAGAACGTCCAGCTGATCAACACCAACGGCAGTTGG  
TCGGAGGAGCGTGGTCCGCGGTTCTGCTTGCAGGTGCTAGTTGTGGTTGCCGTCAACC
- 1261 HisLeuAsnSerThrAlaLeuAsnCysAsnAspSerLeuAsnThrGlyTrpLeuAlaGly  
CACCTCAATAGCACGGCCCTGAACGTCAATGATAGCTCAACACCGGCTGGTTGGCAGGG  
GTGGAGTTATCGTGCCGGGACTTGACGTTACTATCGGAGTTGTGGCCGACCAACCGTCCC
- 1321 LeuPheTyrHisHisLysPheAsnSerSerGlyCysProGluArgLeuAlaSerCysArg  
CTTTCTATCACCACAAGTTCAACTCTTCAGGCTGTCTGAGAGGCTAGCCAGCTGCCGA  
GAAAAGATAGTGGTGTTCAGTTGAGAAGTCCGACAGGACTCTCCGATCGGTGACGGCT
- 1381 ProLeuThrAspPheAspGlnGlyTrpGlyProIleSerTyrAlaAsnGlySerGlyPro  
CCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATCAGTTATGCCAACGGAAGCGGCCCC  
GGGAATGGCTAAACCTGGTCCCGACCCCGGATAGTCAATACGGTTGCCTTCGCCGGGG
- 1441 AspGlnArgProTyrCysTrpHisTyrProProLysProCysGlyIleValProAlaLys  
GACCAGCGCCCTACTGCTGGCACTACCCCCAAAACCTTGCGGTATTGTGCCCGCGAAG  
CTGGTCCGGGGATGACGACCGTATGGGGGGTTTGGAAACGCCATAACAGGGCGCTTC
- 1501 SerValCysGlyProValTyrCysPheThrProSerProValValValGlyThrThrAsp  
AGTGTGTGGTCCGGTATATTGCTTCACTCCAGCCCCGTGGTGGTGGGAACGACCGAC  
TCACACACACAGGCCATATAACGAAGTGAGGGTCGGGGCACCACCACCTTGCTGGCTG
- 1561 ArgSerGlyAlaProThrTyrSerTrpGlyGluAsnAspThrAspValPheValLeuAsn  
AGGTCCGGGCGCGCCACCTACAGCTGGGGTGAAATGATACGGACGTCTTCGTCTTAAC  
TCCAGCCCCGCGGGTGGATGTGACCCCACTTTTACTATGCCTGCAGAAGCAGGAATTG
- 1621 AsnThrArgProProLeuGlyAsnTrpPheGlyCysThrTrpMetAsnSerThrGlyPhe  
AATACCAGGCCACCGCTGGGCAATTGGTTCCGGTTGTACCTGGATGAACCTCACTGGATTC  
TTATGGTCCGGTGGCGACCCGTTAACCAAGCCAACATGGACCTACTTGAGTTGACCTAAG

SUBSTITUTE SHEET

27 / 40

## FIG. 17-3

- 1681 ThrLysValCysGlyAlaProProCysValIleGlyGlyAlaGlyAsnAsnThrLeuHis  
ACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGAGGGGCGGGCAACAACCCCTGCAC  
TGGTTTCACACGCCTCGCGGAGGAACACAGTAGCCTCCCCGCCCCGTTGTTGTGGGACGTG
- 1741 CysProThrAspCysPheArgLysHisProAspAlaThrTyrSerArgCysGlySerGly  
TGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCCACATACTCTCGGTGCGGCTCCGGT  
ACGGGGTGACTAACGAAGGCGTTCTGAGGCCCTGCGGTGTATGAGAGCCACGCCGAGGCCA
- Leu
- 1801 ProTrpIleThrProArgCysLeuValAspTyrProTyrArgLeuTrpHisTyrProCys  
CCCTGGATCACACCCAGGTGCCTGGTTCGACTACCCGTATAGGCTTTGGCATTATCCTTGT  
GGGACCTAGTGTGGGTCCACGGACAGCTGATGGGCATATCCGAAACCGTAATAGGAACA
- 1861 ThrIleAsnTyrThrIlePheLysIleArgMetTyrValGlyGlyValGluHisArgLeu  
ACCATCAACTACACCATATTTAAATCAGGATGTACGTGGGAGGGGTGCAACACAGGCTG  
TGGTAGTTGATGTGGTATAAATTTTAGTCTACATGCACCCCTCCCCAGCTTGTGTCCGAC
- 1921 GluAlaAlaCysAsnTrpThrArgGlyGluArgCysAspLeuGluAspArgAspArgSer  
GAAGCTGCCTGCAACTGGACGCGGGGCGAACGTGCGATCTGGAAGACAGGGACAGGTCC  
CTTCGACGGACGTTGACCTGCGCCCCGCTTGCAACGCTAGACCTTCTGTCCCTGTCCAGG
- 1981 GluLeuSerProLeuLeuLeuThrThrThrGlnTrpGlnValLeuProCysSerPheThr  
GAGCTCAGCCCCGTTACTGCTGACCACTACACAGTGGCAGGTCTCCCGTGTTCCTTCACA  
CTCGAGTCGGGCAATGACGACTGGTGTATGTCACCGTCCAGGAGGCACAAGGAAGTGT
- 2041 ThrLeuProAlaLeuSerThrGlyLeuIleHisLeuHisGlnAsnIleValAspValGln  
ACCCTACCAGCCTTGTCCACCGGCTCATCCACCTCCACCAGAACATGTGGACGTGCAG  
TGGGATGGTTCGGAACAGGTGGCCGGAGTAGGTGGAGGTGGTCTTGTAAACCTGCACGTC
- 2101 TyrLeuTyrGlyValGlySerSerIleAlaSerTrpAlaIleLysTrpGluTyrValVal  
TACTTGTACGGGGTGGGGTCAAGCATCGCGTCTCTGGGCCATTAAGTGGGAGTACGTCGTT  
ATGAACATGCCCCACCCAGTTCGTAGCGCAGGACCCGGTAATTCACCCCTCATGCAGCAA
- 2161 LeuLeuPheLeuLeuLeuAlaAspAlaArgValCysSerCysLeuTrpMetMetLeuLeu  
CTCCTGTTCTCTTCTGCTTGCAGACGCGCGCTCTGCTCCTGCTTGTGGATGATGCTACTC  
GAGGACAAGGAAGACGAACGTCTGCGCGCAGACGAGGACGAACACCTACTACGATGAG
- 2221 IleSerGlnAlaGluAlaAlaLeuGluAsnLeuValIleLeuAsnAlaAlaSerLeuAla  
ATATCCCAAGCGGAGGCGGCTTGGAGAACCTCGTAATATTAATGCAGCATCCCTGGCC  
TATAGGGTTTCGCTCCGCCGAAACCTCTTGGAGCATTATGAATTACGTCGTAGGGACCGG
- 2281 GlyThrHisGlyLeuValSerPheLeuValPhePheCysPheAlaTrpTyrLeuLysGly  
GGGACGCACGGTCTTGTATCCTTCTCTGTTCTTCTGCTTTCATGTTATTTGAAGGGT  
CCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAGACGAAACGTACCATAAACTTCCCA
- 2341 LysTrpValProGlyAlaValTyrThrPheTyrGlyMetTrpProLeuLeuLeuLeuLeu  
AAGTGGGTGCCCCGAGCGGTCTACACCTTCTACGGGATGTGGCCTCTCCTCTGCTCCTG  
TTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCCTACACCGGAGAGGAGGACGAGGAC
- 2401 LeuAlaLeuProGlnArgAlaTyrAlaLeuAspThrGluValAlaAlaSerCysGlyGly  
TTGGCGTTGCCCCAGCGGGCGTACGCGCTGGACACGGAGGTGGCCGCGTCGTGTGGCGGT  
AACCGCAACGGGGTTCGCCCCATGCGCGACCTGTGCTCCACCGGCGCAGCACACCGCCA
- 2461 ValValLeuValGlyLeuMetAlaLeuThrLeuSerProTyrTyrLysArgTyrIleSer  
GTTGTTCTCGTCGGGTGATGGCGCTGACTCTGTACCATATTACAAGCGCTATATCAGC  
CAACAAGAGCAGCCCACTACCGCGACTGAGACAGTGGTATAATGTTCCGGATATAGTCG
- Asn
- 2521 TrpCysLeuTrpTrpLeuGlnTyrPheLeuThrArgValGluAlaGlnLeuHisValTrp  
TGGTGCTTGTGGTGGCTTCAGTATTTTCTGACCAGAGTGGAAGCGCACTGCACGTGTGG  
ACCACGAACACCACCGAAGTCATAAAAGACTGGTCTCACCTTCGCGTTGACGTGCACACC
- 2581 IleProProLeuAsnValArgGlyGlyArgAspAlaValIleLeuLeuMetCysAlaVal  
ATTCCCCCCTCAACGTCCGAGGGGGCGGACGCGCTCATCTTACTCATGTGTGCTGTA  
TAAGGGGGGGAGTTGCAGGCTCCCCCGCGCTGCGGCAGTAGAATGAGTACACACGACAT

SUBSTITUTE SHEET

28 / 40

## FIG. 17-4

2641 HisProThrLeuValPheAspIleThrLysLeuLeuAlaValPheGlyProLeuTrp  
CACCCGACTCTGGTATTTGACATCACCAAATTGCTGCTGGCCGCTCTCGGACCCCTTTGG  
GTGGGCTGAGACCATAAACTGTAGTGGTTTAAACGACGACCGGCAGAACCTGGGGAAACC

2701 IleLeuGlnAlaSerLeuLeuLysValProTyrPheValArgValGlnGlyLeuLeuArg  
ATTCTTCAAGCCAGTTTGTCTAAAGTACCCTACTTTGTGCGCGTCCAAGGCCTTCTCCGG  
TAAGAAGTTCGGTCAAACGAATTTTCATGGGATGAAACACGCGCAGGTTCGGAAGAGGCC

2761 PheCysAlaLeuAlaArgLysMetIleGlyGlyHisTyrValGlnMetValIleIleLys  
TTCTGCGCGTTAGCGCGGAAGATGATCGGAGGCCATTACGTGCAATGGTCATCATTAAG  
AAGACGCGCAATCGCGCCTTCTACTAGCCTCCGGTAATGCAGTTTACCAGTAGTAATTC

2821 LeuGlyAlaLeuThrGlyThrTyrValTyrAsnHisLeuThrProLeuArgAspTrpAla  
TTAGGGGCGCTTACTGGCACCTATGTTTATAACCATCTCACTCCTCTTCGGGACTGGGCG  
AATCCCCGGAATGACCGTGGATACAAATATTGGTAGAGTGAGGAGAAGCCCTGACCCGC

2881 HisAsnGlyLeuArgAspLeuAlaValAlaValGluProValValPheSerGlnMetGlu  
CACAACGGCTTGCAGATCTGGCCGTGGCTGTAGAGCCAGTCGTCTTCTCCCAAATGGAG  
GTGTTGCCGAACGCTCTAGACCGGCACCGACATCTCGGTCAGCAGAAGAGGGTTTACCTC

2941 ThrLysLeuIleThrTrpGlyAlaAspThrAlaAlaCysGlyAspIleIleAsnGlyLeu  
ACCAAGCTCATCAGTGGGGGGCAGATACCGCCGCGTGGGTGACATCATCAACGGCTTG  
TGGTTCGAGTAGTGACCCCCGCTCTATGGCGGCGCACGCCACTGTAGTAGTTGCCGAAC

3001 ProValSerAlaArgArgGlyArgGluIleLeuLeuGlyProAlaAspGlyMetValSer  
CCTGTTTCCGCCCCGAGGGGCGGGAGATACTGCTCGGGCCAGCCGATGGAATGGTCTCC  
GGACAAAGGCGGGCGTCCCCGGCCCTCTATGACGAGCCGCTCGGCTACCTTACCAGAGG

3061 LysGlyTrpArgLeuLeuAlaProIleThrAlaTyrAlaGlnGlnThrArgGlyLeuLeu  
AAGGGGTGGAGGTTGTGCGGCCCATCACGGCGTACGCCCAGCAGACAAGGGGCCTCCTA  
TCCCCACCTCCAACGACCGCGGGTAGTGCCGCATGCGGGTCTGTTCCTCCCGGAGGAT

3121 GlyCysIleIleThrSerLeuThrGlyArgAspLysAsnGlnValGluGlyGluValGln  
GGGTGCATAATCACCAGCCTAACTGGCCGGGACAAAAACCAAGTGGAGGTGAGGTCCAG  
CCCACGTATTAGTGTGCGATTGACCGGCCCTGTTTTTGGTTACCTCCCACTCCAGGTC

3181 IleValSerThrAlaAlaGlnThrPheLeuAlaThrCysIleAsnGlyValCysTrpThr  
ATTGTGTCAACTGCTGCCAAACCTTCTGCGCAACGTGCATCAATGGGGTGTGTGCGACT  
TAACACAGTTGACGACGGGTTTGAAGGACCGTTGCACGTAGTTACCCACACGACCTGA

3241 ValTyrHisGlyAlaGlyThrArgThrIleAlaSerProLysGlyProValIleGlnMet  
GTCTACCACGGGGCCGGAACGAGGACCATCGCGTCACCCAAGGTCCTGTGCATCCAGATG  
CAGATGGTGCCCCGGCCTTGCTCCTGGTAGCGCAGTGGGTTCCAGGACAGTAGGTCTAC

3301 TyrThrAsnValAspGlnAspLeuValGlyTrpProAlaProGlnGlySerArgSerLeu  
TATACCAATGTAGACCAAGACCTTGTGGGCTGGCCCGCTCCGCAAGGTAGCCGCTCATTG  
ATATGGTTACATCTGGTTCTGGAACACCCGACCGGGCGAGGCGTTCCATCGGCGAGTAAC

3361 ThrProCysThrCysGlySerSerAspLeuTyrLeuValThrArgHisAlaAspValIle  
ACACCCTGCACTTGC GGCTCCTCGGACCTTTACCTGGTCACGAGGCACGCCGATGTCATT  
TGTGGGACGTGAACGCGGAGGAGCTGGAAATGGACAGTGTCCGTGCGGCTACAGTAA

3421 ProValArgArgArgGlyAspSerArgGlySerLeuLeuSerProArgProIleSerTyr  
CCCGTGC GCGCGGGGTGATAGAGGGGCGAGCCTGCTGTGCCCCGCGCCCATTTCCCTAC  
GGGCACGCGCGCGCCCACTATCGTCCCCGTGCGACGACAGCGGGGCGGGGTAAAGGATG

3481 LeuLysGlySerSerGlyGlyProLeuLeuCysProAlaGlyHisAlaValGlyIlePhe  
TTGAAAGGCTCCTCGGGGGGTCCGCTGTGTGCCCCGCGGGGCACGCCGTGGGCATATTT  
AACTTTCCGAGGAGCCCCCAGGCGACAACACGGGGCGCCCCGTGCGGCACCCGTATAAA

3541 ArgAlaAlaValCysThrArgGlyValAlaLysAlaValAspPheIleProValGluAsn  
AGGGCCGCGGTGTGCACCCGTGGAGTGGCTAAGGCGGTGGACTTTATCCCTGTGGAGAAC  
TCCCGCGCCACACGTGGGCACCTACCGATTCCGCCACCTGAAATAGGGACACCTCTTG

SUBSTITUTE SHEET

29/40

FIG. 17-5

- 3601 LeuGluThrThrMetArgSerProValPheThrAspAsnSerSerProProValValPro  
CTAGAGACAACCATGAGGTCCCCGGTGTTCACGGATAACTCCTCTCCACCAGTAGTGCCC  
GATCTCTGTTGGTACTCCAGGGGCCACAAGTGCCTATTGAGGAGAGGTGGTCATCAGGGG
- 3661 GlnSerPheGlnValAlaHisLeuHisAlaProThrGlySerGlyLysSerThrLysVal  
CAGAGCTTCCAGGTGGCTCACCTCCATGCTCCACAGGCAGCGCAAAGCACCAAGGTC  
GTCTCGAAGGTCCACCGAGTGGAGGTACGAGGGTGTCCGTCGCCGTTTTCGTGGTTCCAG
- 3721 ProAlaAlaTyrAlaAlaGlnGlyTyrLysValLeuValLeuAsnProSerValAlaAla  
CCGGCTGCATATGCAGCTCAGGGCTATAAGGTGCTAGTACTCAACCCCTCTGTTGCTGCA  
GGCCGACGTATACGTGAGTCCCGATATTCCACGATCATGAGTTGGGGAGACAACGACGT
- 3781 ThrLeuGlyPheGlyAlaTyrMetSerLysAlaHisGlyIleAspProAsnIleArgThr  
ACACTGGGCTTTGGTGCTTACATGTCCAAGGCTCATGGGATCGATCCTAACATCAGGACC  
TGTGACCCGAAACCACGAATGTACAGGTTCCGAGTACCCTAGCTAGGATTGTAGTCTGG
- 3841 GlyValArgThrIleThrThrGlySerProIleThrTyrSerThrTyrGlyLysPheLeu  
GGGGTGAGAACAATTACCACTGGCAGCCCCATCAGTACTCCACCTACGGCAAGTTCCTT  
CCCCACTCTTGTAAATGGTGACCGTCGGGGTAGTGCATGAGGTGGATGCCGTTCAAGGAA
- 3901 AlaAspGlyGlyCysSerGlyGlyAlaTyrAspIleIleIleCysAspGluCysHisSer  
GCCGACGGCGGGTGCTCGGGGGGCGCTTATGACATAATAATTTGTGACGAGTGCCACTCC  
CGGCTGCCGCCCCACGAGCCCCCGGAATACTGTATTATTAAACACTGCTCAGGTTGAGG
- 3961 ThrAspAlaThrSerIleLeuGlyIleGlyThrValLeuAspGlnAlaGluThrAlaGly  
ACGGATGCCACATCCATCTTGGGCATCGGCACTGTCTTGACCAAGCAGAGACTGCCGGG  
TGCTTACGGTGTAGGTAGAACCCGTAGCCGTGACAGGAAGTGGTTCGTCTCTGACGCCCC
- 4021 AlaArgLeuValValLeuAlaThrAlaThrProProGlySerValThrValProHisPro  
GCGAGACTGGTTGTGCTCGCCACCGCCACCCCTCCGGGCTCCGTCAGTGTGCCCCATCCC  
CGCTCTGACCAACACGAGCGGTGGCGGTGGGGAGGCCCGAGGCAGTGACAGGGGTAGGG
- 4081 AsnIleGluGluValAlaLeuSerThrThrGlyGluIleProPheTyrGlyLysAlaIle  
AACATCGAGGAGGTGCTCTGTCCACCACCGGAGAGATCCCTTTTACGGCAAGGCTATC  
TTGTAGCTCCTCCAACGAGACAGGTGGTGGCCTCTCTAGGGAAAATGCCGTTCCGATAG
- 4141 ProLeuGluValIleLysGlyGlyArgHisLeuIlePheCysHisSerLysLysLysCys  
CCCCTCGAAGTAATCAAGGGGGGAGACATCTCATCTTCTGTCAATCAAGAAGAAGTGC  
GGGGAGCTTCATTAGTTCCCCCCTCTGTAGAGTAGAAGACAGTAAGTTTCTTCTTCACG
- 4201 AspGluLeuAlaAlaLysLeuValAlaLeuGlyIleAsnAlaValAlaTyrTyrArgGly  
GACGAACCTCGCCGCAAAGCTGGTCGATTGGGCATCAATGCCGTGGCTACTACCGCGGT  
CTGCTTGAGCGGCGTTTCGACCAGCGTAACCCGTAGTTACGGCACCGGATGATGGCGCCA
- 4261 LeuAspValSerValIleProThrSerGlyAspValValValAlaThrAspAlaLeu  
CTTGACGTGTCCGTATCCCGACCAGCGCGATGTTGTGTCGTGGCAACCGATGCCCTC  
GAACTGCACAGGCAGTAGGGCTGGTCGCCGTACAACAGCAGCACCGTTGGCTACGGGAG
- 4321 MetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsnThrCysValThrGln  
ATGACCGGCTATACCGGCGACTTCGACTCGGTGATAGACTGCAATACGTGTGTACCCAG  
TACTGGCCGATATGGCCGTGAAGCTGAGCCACTATCTGACGTTATGCACACAGTGGGTC
- 4381 ThrValAspPheSerLeuAspProThrPheThrIleGluThrIleThrLeuProGlnAsp  
ACAGTCGATTTTCAGCCTTGACCCTACCTTCACCATTGAGACAATCAGCTCCCCCAGGAT  
TGTCAGCTAAAGTCGGAAGTGGGATGGAAGTGGTAAGTCTGTTAGTGCGAGGGGGTCCTA
- 4441 AlaValSerArgThrGlnArgArgGlyArgThrGlyArgGlyLysProGlyIleTyrArg  
GCTGTCTCCCGCACTCAACGTCGGGGCAGGACTGGCAGGGGGAAGCCAGGCATCTACAGA  
CGACAGAGGGCGTGAGTTGCAGCCCCGTCTGACCGTCCCCCTTCGGTCCGTAGATGTCT
- 4501 PheValAlaProGlyGluArgProSerGlyMetPheAspSerSerValLeuCysGluCys  
TTTGTCGACCGGGGGAGCGCCCTCCGGCATGTTGACTCGTCCGTCTCTGTGAGTGC  
AAACACCGTGGCCCCCTCGCGGGGAGGCCGTACAAGCTGAGCAGGCAGGAGACACTCAGC

SUBSTITUTE SHEET



30 / 40

## FIG. 17-6

4561 TyrAspAlaGlyCysAlaTrpTyrGluLeuThrProAlaGluThrThrValArgLeuArg  
TATGACGCAGGCTGTGCTTGGTATGAGCTCACGCCGCGAGACTACAGTTAGGCTACGA  
ATACTGCGTCCGACACGAACCATACTCGAGTGGGGCGGCTCTGATGTCAATCCGATGCT

4621 AlaTyrMetAsnThrProGlyLeuProValCysGlnAspHisLeuGluPheTrpGluGly  
GCGTACATGAACACCCCGGGGCTTCCCGTGTGCCAGGACCATCTTGAATTTTGGGAGGGC  
CGCATGTACTTGTGGGGCCCCGAAGGGCACACGGTCTGGTAGAACCTAAACCCCTCCCC

4681 ValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThrLysGlnSerGly  
GTCTTTACAGGCCCTCACTCATATAGATGCCCACTTTCTATCCCAGACAAAGCAGAGTGGG  
CAGAAATGTCCGGAGTGAGTATATCTACGGGTGAAAGATAGGGTCTGTTTCGTCTCACCC

4741 GluAsnLeuProTyrLeuValAlaTyrGlnAlaThrValCysAlaArgAlaGlnAlaPro  
GAGAACCTTCCTTACCTGGTAGCGTACCAAGCCACCGTGTGCGCTAGGGCTCAAGCCCCT  
CTCTTGAAGGAATGGACCATCGCATGGTTCGGTGGCACACGGATCCCGAGTTCGGGGG

4801 ProProSerTrpAspGlnMetTrpLysCysLeuIleArgLeuLysProThrLeuHisGly  
CCCCATCGTGGGACCAGATGTGGAAGTGTGTTGATTGCGCTCAAGCCCACCTCCATGGG  
GGGGTAGCACCCCTGGTCTACACCTTCACAACTAAGCGGAGTTCGGGTGGGAGGTACCC

4861 ProThrProLeuLeuTyrArgLeuGlyAlaValGlnAsnGluIleThrLeuThrHisPro  
CCAACACCCCTGCTATACAGACTGGGCGCTGTTCAGAATGAAATCACCTGACGCACCCA  
GGTTGTGGGGACGATATGTCTGACCCGCGACAAGTCTTACTTTAGTGGGACTGCGTGGGT

4921 ValThrLysTyrIleMetThrCysMetSerAlaAspLeuGluValValThrSerThrTrp  
GTCACCAAATACATCATGACATGCATGTGCGCCGACCTGGAGGTGCTCAGCAGCACCTGG  
CAGTGGTTTATGTAGTACTGTACGTACAGCCGGCTGGACCTCCAGCAGTGCCTCGTGGACC

4981 ValLeuValGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSerThrGlyCysVal  
GTGCTCGTTGGCGGCGTCTGGCTGCTTTGGCCGCGTATTGCCTGTCAACAGCTGCGGTG  
CACGAGCAACCGCCGAGGACCGACGAAACCGGCGCATAACGGACAGTTGTCCGACGCAC

5041 ValIleValGlyArgValValLeuSerGlyLysProAlaIleIleProAspArgGluVal  
GTCATAGTGGGCAGGGTCGTCTTGTCCGGGAAGCCGGAATCATACCTGACAGGGAAGTC  
CAGTATCACCCGTCCAGCAGAACAGGCCCTTCGGCCGTTAGTATGGACTGTCCCTTCAG

5101 LeuTyrArgGluPheAspGluMetGluGluCysSerGlnHisLeuProTyrIleGluGln  
CTCTACCGAGAGTTCGATGAGATGGAAGAGTGCTCTCAGCACTTACCGTACATCGAGCAA  
GAGATGGCTCTCAAGCTACTCTACCTTCTCACGAGAGTCGTGAATGGCATGTAGCTCGTT

5161 GlyMetMetLeuAlaGluGlnPheLysGlnLysAlaLeuGlyLeuLeuGlnThrAlaSer  
GGGATGATGCTCGCCGAGCAGTTCAAGCAGAAGGCCCTCGGCCTCTGCAGACCGCGTCC  
CCCTACTACGAGCGGCTCGTCAAGTTCGTCTTCCGGGAGCCGGAGGACGTCTGGCGCAGG

5221 ArgGlnAlaGluValIleAlaProAlaValGlnThrAsnTrpGlnLysLeuGluThrPhe  
CGTCAGGCAGAGGTATCGCCCCCTGCTGCCAGACCAACTGGCAAACTCGAGACCTTC  
GCAGTCCGTCTCCAATAGCGGGGACGACAGGTCTGGTTGACCGTTTTTGAGCTCTGGAAG

5281 TrpAlaLysHisMetTrpAsnPheIleSerGlyIleGlnTyrLeuAlaGlyLeuSerThr  
TGGGCGAAGCATATGTGGAACCTCATCAGTGGGATACAATACTTGGCGGGCTTGTCAACG  
ACCCGCTTCGTATACACCTTGAAGTAGTCACCTATGTTATGAACGCCCCGAACAGTTGC

5341 LeuProGlyAsnProAlaIleAlaSerLeuMetAlaPheThrAlaAlaValThrSerPro  
CTGCCTGGTAACCCCGCCATTGCTTCATTGATGGCTTTTACAGCTGCTGTACACGCCCA  
GACGGACCATTGGGGCGGTAACGAAGTAACACCGAAATGTCGACGACAGTGGTCGGGT

5401 LeuThrThrSerGlnThrLeuLeuPheAsnIleLeuGlyGlyTrpValAlaAlaGlnLeu  
CTAACCACTAGCCAAACCTCCTTCAACATATTGGGGGGGTGGGTGGCTGCCAGCTC  
GATTGGTGATCGGTTTGGGAGGAGAAGTTGTATAACCCCCACCCACCGACGGGTGCGAG

5461 AlaAlaProGlyAlaAlaThrAlaPheValGlyAlaGlyLeuAlaGlyAlaAlaIleGly  
GCCGCCCCGGTGCCGCTACTGCCTTGTGGGCGCTGGCTTAGCTGGCGCCGCCATCGGC  
CGGCGGGGCCACGGCGATGACGGAACACCCGCGACCGAATCGACCGCGGCGGTAGCCG

SerValGlyLeuGlyLysValLeuIleAspIleLeuAlaGlyTyrGlyAlaGlyValAla

SUBSTITUTE SHEET

31 / 40

## FIG. 17-7

- 5521 AGTGTGGACTGGGGAAGGTCCTCATAGACATCCTTGCAGGGTATGGCGCGGGCGTGGCG  
TCACAACCTGACCCCTTCCAGGAGTATCTGTAGGAACGTCCCATACCGCGCCGACCCG
- Gly  
5581 GlyAlaLeuValAlaPheLysIleMetSerGlyGluValProSerThrGluAspLeuVal  
GGAGCTCTTGTGGCATTCAGATCATGAGCGGTGAGGTCCCCTCCACGGAGGACCTGGTC  
CCTCGAGAACACCGTAAGTTCTAGTACTCGCCACTCCAGGGGAGGTGCCTCTGGACCAG
- 5641 AsnLeuLeuProAlaIleLeuSerProGlyAlaLeuValValGlyValValCysAlaAla  
AATCTACTGCCCCCATCTCTCGCCCGGAGCCCTCGTAGTCGGCGTGGTCTGTGCAGCA  
TTAGATGACGGGCGGTAGGAGAGCGGGCCTCGGGAGCATCAGCCGACCCAGACACGTCGT
- 5701 IleLeuArgArgHisValGlyProGlyGluGlyAlaValGlnTrpMetAsnArgLeuIle  
ATACTGCGCCGGCACGTGGCCCGGGCGAGGGGCGAGTGCAGTGGATGAACCGGCTGATA  
TATGACGCGGCCGTGCAACCGGGCCCGTCCCCCGTCACGTCACCTACTTGGCCGACTAT
- 5761 AlaPheAlaSerArgGlyAsnHisValSerProThrHisTyrValProGluSerAspAla  
GCCTTCGCCTCCCGGGGAACCATGTTTCCCCACGCACTACGTGCCGGAGAGCGATGCA  
CGGAAGCGAGGGCCCCCTTGGTACAAAGGGGTGCGTGATGCACGGCCTCTCGCTACGT
- HisCys  
5821 AlaAlaArgValThrAlaIleLeuSerSerLeuThrValThrGlnLeuLeuArgArgLeu  
GCTGCCCGCGTCACTGCCATACTCAGCAGCCTCACTGTAACCCAGCTCCTGAGGCGACTG  
CGACGGGCGCAGTGACGGTATGAGTCGTGAGGTGACATTGGGTGCGAGGACTCCGCTGAC
- 5881 HisGlnTrpIleSerSerGluCysThrThrProCysSerGlySerTrpLeuArgAspIle  
CACCAGTGGATAAGCTCGGAGTGTAACCTCCATGCTCCGGTTCCTGGCTAAGGGACATC  
GTGGTCACCTATTTCGAGCCTCACATGGTGAGGTACGAGGCCAAGGACCGATTCCCTGTAG
- 5941 TrpAspTrpIleCysGluValLeuSerAspPheLysThrTrpLeuLysAlaLysLeuMet  
TGGGACTGGATATGCGAGGTGTTGAGCGACTTTAAGACCTGGCTAAAAGCTAAGCTCATG  
ACCCTGACCTATACGCTCCACAACCTCGCTGAAATTCTGGACCGATTTTCGATTTCGAGTAC
- 6001 ProGlnLeuProGlyIleProPheValSerCysGlnArgGlyTyrLysGlyValTrpArg  
CCACAGCTGCCTGGGATCCCCCTTGTGTCTCTGCCAGCGCGGGTATAAGGGGGTCTGGCGA  
GGTGTGCGACGGACCCTAGGGGAACACAGGACGGTTCGCGCCCATATTCCCCCAGACCGCT
- Gly  
6061 ValAspGlyIleMetHisThrArgCysHisCysGlyAlaGluIleThrGlyHisValLys  
GTGGACGGCATCATGCACACTCGCTGCCACTGTGGAGCTGAGATCACTGGACATGTCAAA  
CACCTGCCGTAGTACGTGTGAGCGACGGTGACACCTCGACTCTAGTGACCTGTACAGTTT
- 6121 AsnGlyThrMetArgIleValGlyProArgThrCysArgAsnMetTrpSerGlyThrPhe  
AACGGGACGATGAGGATCGTCCGGTCTAGGACCTGCAGGAACATGTGGAGTGGGACCTTC  
TTGCCCTGCTACTCTAGCAGCCAGGATCCTGGACGTCTTGTACACCTCACCTTGAAG
- 6181 ProIleAsnAlaTyrThrThrGlyProCysThrProLeuProAlaProAsnTyrThrPhe  
CCCATTAAATGCCTACACCACGGGCCCCCTGTACCCCCCTTCTGCGCCGAACCTACAGTTT  
GGGTAATTACGGATGTGGTGCCCGGGGACATGGGGGAAGGACGGGCTTGATGTGCAAG
- 6241 AlaLeuTrpArgValSerAlaGluGluTyrValGluIleArgGlnValGlyAspPheHis  
GCGCTATGGAGGGTGTCTGCAGAGGAATATGTGGAGATAAGGCAGGTGGGGGACTTCCAC  
CGCGATACCTCCACAGACGTCTCCTTATACACCTCTATTCCGTCCACCCCTGAAGGTG
- 6301 TyrValThrGlyMetThrThrAspAsnLeuLysCysProCysGlnValProSerProGlu  
TACGTGACGGGTATGACTACTGACAATCTCAAATGCCCGTGCCAGGTCCCATCGCCCGAA  
ATGCACTGCCCATACTGATGACTGTTAGAGTTTACGGGCACGGTCCAGGGTAGCGGGCTT
- 6361 PhePheThrGluLeuAspGlyValArgLeuHisArgPheAlaProProCysLysProLeu  
TTTTTCACAGAATTGGACGGGGTGCCTTACATAGGTTTGGCGCCCCCTGCAAGCCCTTG  
AAAAAGTGTCTTAACCTGCCCCACGGGATGTATCAAACCGGGGGGACGTTTCGGGAAC
- 6421 LeuArgGluGluValSerPheArgValGlyLeuHisGluTyrProValGlySerGlnLeu  
CTGCGGGAGGAGGTATCATTCAGAGTAGGACTCCACGAATACCCGGTAGGGTTCGCAATTA  
GACGCCCTCCTCCATAGTAAGTCTCATCTGAGGTGCTTATGGGCCATCCAGCGTTAAT

SUBSTITUTE SHEET

32 / 40

## FIG. 17-8

6481 ProCysGluProGluProAspValAlaValLeuThrSerMetLeuThrAspProSerHis  
CCTTGCGAGCCCGAACCGGACGTGGCCGTGTTGACGTCCATGCTCACTGATCCCTCCCAT  
GGAACGCTCGGGCTTGGCCTGCACCGGCACAACGTCAGGTACGAGTGACTAGGGAGGGTA

6541 IleThrAlaGluAlaAlaGlyArgArgLeuAlaArgGlySerProProSerValAlaSer  
ATAACAGCAGAGCGCGCCGGCGAAGGTGGCGAGGGGATCACCCCTCTGTGGCCAGC  
TATTGTCGTCTCCGCCGCGCTTCCAACCGCTCCCTAGTGGGGGAGACACCGGTG

6601 SerSerAlaSerGlnLeuSerAlaProSerLeuLysAlaThrCysThrAlaAsnHisAsp  
TCCTCGGCTAGCCAGCTATCCGCTCCATCTCTCAAGGCAACTTGCACCGTAACCATGAC  
AGGAGCCGATCGGTGATAGGCGAGGTAGAGAGTTCCGTTGAACGTGGCGATTGGTACTG

6661 SerProAspAlaGluLeuIleGluAlaAsnLeuLeuTrpArgGlnGluMetGlyGlyAsn  
TCCCTGTAGTCTGAGCTCATAGAGGCCAACCTCCTATGGAGGCAGGAGATGGGCGGCAAC  
AGGGGACTACGACTCGAGTATCTCCGGTGGAGGATACCTCCGTCTCTACCCGCCGTTG

6721 IleThrArgValGluSerGluAsnLysValValIleLeuAspSerPheAspProLeuVal  
ATCACCAGGGTTGAGTCAGAAAACAAAGTGGTGATTCTGGACTCCTTCGATCCGCTTGTG  
TAGTGGTCCCAACTCAGTCTTTTGTTCACCACTAAGACCTGAGGAAGCTAGGCGAACAC

6781 AlaGluGluAspGluArgGluIleSerValProAlaGluIleLeuArgLysSerArgArg  
GCGAGGAGGACGAGCGGGAGATCTCCGTACCCGCAGAAATCTGCGGAAGTCTCGGAGA  
CGCTCCTCTGCTCGCCCTCTAGAGGCATGGGCGTCTTTAGGACGCTTCAGAGCCTCT

6841 PheAlaGlnAlaLeuProValTrpAlaArgProAspTyrAsnProProLeuValGluThr  
TTCGCCCAGGCCCTGCCCGTTTGGGCGCGGCCGACTATAACCCCGCTAGTGGAGACG  
AAGCGGGTCCGGGACGGGCAAACCCGCGCGCTGATATTGGGGGGCGATCACCTCTGC

6901 TrpLysLysProAspTyrGluProProValValHisGlyCysProLeuProProProLys  
TGGAAAAGCCCGACTACGAACCACCTGTGGTCCATGGCTGTCCGCTTCCACCTCCAAAG  
ACCTTTTTTCGGGCTGATGCTTGGTGGACACCAAGTACCGACAGGCGAAGGTGGAGGTTTC

6961 SerProProValProProProArgLysLysArgThrValValLeuThrGluSerThrLeu  
TCCCTCCTGTGCTCCGCTCGGAAGAAGCGGACGGTGGTCTCACTGAATCAACCCTA  
AGGGGAGGACACGGAGGCGAGCCTTCTTCGCTGCCACCAGGAGTGACTTAGTTGGGAT

7021 Ser  
SerThrAlaLeuAlaGluLeuAlaThrArgSerPheGlySerSerSerThrSerGlyIle  
TCTACTGCCTTGGCCGAGCTCGCCACCAGAAGCTTTGGCAGCTCCTCACTTCCGGCATT  
AGATGACGGAACCGGCTCGAGCGGTGGTCTTCGAAACCGTCGAGGAGTTGAAGCCGTAA

7081 ThrGlyAspAsnThrThrThrSerSerGluProAlaProSerGlyCysProProAspSer  
ACGGGCGACAATACGACAACATCCTCTGAGCCCGCCCTTCTGGCTGCCCGCCGACTCC  
TGCCCGCTGTTATGCTTGTAGGAGACTCGGGCGGGGAAGACCGACGGGGGGGCTGAGG

7141 PheAla  
AspAlaGluSerTyrSerSerMetProProLeuGluGlyGluProGlyAspProAspLeu  
GACGCTGAGTCCTATTCTCCATGCCCGCCCTGGAGGGGGAGCCTGGGGATCCGGATCTT  
CTGGCACTCAGGATAAGGAGGTACGGGGGGACCTCCCCCTCGGACCCCTAGGCCTAGAA

7201 SerAspGlySerTrpSerThrValSerSerGluAlaAsnAlaGluAspValValCysCys  
AGCGACGGGTCACTGGTCAACGGTCAGTAGTGAGGCCAACGCGGAGGATGTCGTGTGCTGC  
TCGCTGCCAGTACCAGTTGCCAGTCATCACTCCGGTTGCGCTCCTACAGCACACGACG

7261 SerMetSerTyrSerTrpThrGlyAlaLeuValThrProCysAlaAlaGluGluGlnLys  
TCAATGTCTTACTCTTGGACAGGCGCACTCGTCACCCCGTGGCGCGGAAGAAGAGAAA  
AGTTACAGAATGAGAACCTGTCCGCTGAGCAGTGGGGCACGCGCGCCTTCTTGTCTTT

7321 LeuProIleAsnAlaLeuSerAsnSerLeuLeuArgHisHisAsnLeuValTyrSerThr  
CTGCCCATCAATGCACTAAGCACTCGTTGCTACGTACCAATTTGGTGTATTCCACC  
GACGGGTAGTTACGTGATTCTGTGAGCAACGATGCAGTGGTGTAAACCATAGGTGG

7381 ThrSerArgSerAlaCysGlnArgGlnLysLysValThrPheAspArgLeuGlnValLeu  
ACCTCACGCAGTGCTTGCCAAAGGCAGAAAGTCACATTTGACAGACTGCAAGTTCTG  
TGGAGTGGCTCACGAACGGTTTCCGTCTTCTTTTCAAGTGTAACTGTCTGACGTTCAAGAC

SUBSTITUTE SHEET

33 / 40

## FIG. 17-9

- 7441 AspSerHisTyrGlnAspValLeuLysGluValLysAlaAlaAlaSerLysValLysAla  
GACAGCCATTACCAGGACGTACTCAAGGAGGTTAAAGCAGCGCGTCAAAAGTGAAGGCT  
CTGTCGGTAATGGTCCTGCATGAGTTCTTCCAATTCGTCGCCGAGTTTTCACCTCCGA
- Phe
- 7501 AsnLeuLeuSerValGluGluAlaCysSerLeuThrProProHisSerAlaLysSerLys  
AACTTGCTATCCGTAGAGGAGCTTGCAGCCTGACGCCCCACACTCAGCCAAATCCAAG  
TTGAACGATAGGCATCTCCTTCGAACGTCGGACTGCGGGGTGTGAGTCGGTTTAGGTTT
- 7561 PheGlyTyrGlyAlaLysAspValArgCysHisAlaArgLysAlaValThrHisIleAsn  
TTTGGTTATGGGGCAAAGACGTCCGTTGCCATGCCAGAAAGGCCGTAACCCACATCAAC  
AAACCAATACCCCGTTTCTGCAGGCAACGGTACGGTCTTCCGGCATTTGGGTGTAGTTG
- 7621 SerValTrpLysAspLeuLeuGluAspAsnValThrProIleAspThrThrIleMetAla  
TCCGTGTGGAAGACCTTCTGGAAGACAATGTAAACCAATAGACACTACCATCATGGCT  
AGGCACACCTTCTGGAAGACCTTCTGTTACATTGTGGTTATCTGTGATGGTAGTACCGA
- 7681 LysAsnGluValPheCysValGlnProGluLysGlyGlyArgLysProAlaArgLeuIle  
AAGAACGAGGTTTCTGCGTTTACGCTGAGAAGGGGGTTCGTAAGCCAGCTCGTCTCATC  
TTCTTGCTCCAAAGACGCAAGTCGGACTCTTCCCCCAGCATTGGTTCGAGCAGAGTAG
- 7741 ValPheProAspLeuGlyValArgValCysGluLysMetAlaLeuTyrAspValValThr  
GTGTTCCCCGATCTGGGCGTGCCTGTGCGAAAAGATGGCTTTGTACGACGTGGTTACA  
CACAAGGGGCTAGACCCGCACGCGCACACGCTTTTCTACCGAAACATGCTGCACCAATGT
- 7801 LysLeuProLeuAlaValMetGlySerSerTyrGlyPheGlnTyrSerProGlyGlnArg  
AAGCTCCCTTGGCCGTGATGGGAAGCTCCTACGGATTCCAATACTCACCAGGACAGCGG  
TTCGAGGGGAACCGGCACTACCTTCGAGGATGCTAAGGTTATGAGTGGTCTGTGCGC
- 7861 ValGluPheLeuValGlnAlaTrpLysSerLysLysThrProMetGlyPheSerTyrAsp  
GTTGAATTCTCGTGCAAGCGTGAAGTCCAAGAAAACCCCAATGGGGTTCTCGTATGAT  
CACTTAAGGAGCACGTTTCGCACCTTCAGGTTCTTTTGGGGTTACCCCAAGAGCATACTA
- 7921 ThrArgCysPheAspSerThrValThrGluSerAspIleArgThrGluGluAlaIleTyr  
ACCCGCTGCTTTGACTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTAC  
TGGGCGACGAAACTGAGGTGTCAGTGACTCTCGCTGTAGGCATGCCTCCTCCGTTAGATG
- 7981 GlnCysCysAspLeuAspProGlnAlaArgValAlaIleLysSerLeuThrGluArgLeu  
CAATGTTGTGACCTCGACCCCAAGCCCGTGGCCATCAAGTCCCTCACCAGAGAGGCTT  
GTTACAACACTGGAGCTGGGGGTTCCGGGCGACCGGTAGTTCAGGGAGTGGCTCTCCGAA
- Gly
- 8041 TyrValGlyGlyProLeuThrAsnSerArgGlyGluAsnCysGlyTyrArgArgCysArg  
TATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGGCTATCGCAGGTGCCGC  
ATACAACCCCGGGAATGGTTAAGTTCCCCCTCTTGACGCCGATAGCGTCCACGGCG
- 8101 AlaSerGlyValLeuThrThrSerCysGlyAsnThrLeuThrCysTyrIleLysAlaArg  
GCGAGCGGCGTACTGACAAGTGTGTTAACACCCTCACTTGCTACATCAAGGCCCGG  
CGCTCGCCGATGACTGTTGATCGACACCATTTGTTGGGAGTGAACGATGTAGTTCCGGGCC
- 8161 AlaAlaCysArgAlaAlaGlyLeuGlnAspCysThrMetLeuValCysGlyAspAspLeu  
GCAGCCTGTCGAGCCGAGGGCTCCAGGACTGCACCATGCTCGTGTGTGGCGACGACTTA  
CGTCGGACAGCTCGGCGTCCCGAGGTCTGACGTGGTACGAGCACACCGCTGCTGAAT
- 8221 ValValIleCysGluSerAlaGlyValGlnGluAspAlaAlaSerLeuArgAlaPheThr  
GTCGTTATCTGTGAAAGCGCGGGGGTCCAGGAGGACGCGGCGAGCCTGAGAGCCTTCACG  
CAGCAATAGACACTTTCGCGCCCCCAGGTCTCTGCGCCGCTCGGACTCTCGGAAGTGC
- 8281 GluAlaMetThrArgTyrSerAlaProProGlyAspProProGlnProGluTyrAspLeu  
GAGGCTATGACCAGGTACTCCGCCCCCTGGGGACCCCCACAACCAGAATACGACTTG  
CTCCGATACTGGTCCATGAGGCGGGGGGACCCCTGGGGGGTGTGGTCTTATGCTGAAC
- 8341 GluLeuIleThrSerCysSerSerAsnValSerValAlaHisAspGlyAlaGlyLysArg  
GAGCTCATAACATCATGCTCTCCAACGTGTGAGTCGCCACGACGGCGCTGGAAAGAGG  
CTCGAGTATTGTAGTACGAGGAGGTTGCACAGTCAGCGGGTGTGCGCGACCTTTCTCC

SUBSTITUTE SHEET

34 / 40

8401 ValTyrTyrLeuThrArgAspProThrThrProLeuAlaArgAlaAlaTrpGluThrAla  
GTCTACTACCTCACCCTGACCCCTACAACCCCTCGCGAGAGCTGCGTGGGAGACAGCA  
CAGATGATGGAGTGGGCACCTGGGATGTTGGGGGAGCGCTCTCGACGCACCCTCTGTCGT

8461 ArgHisThrProValAsnSerTrpLeuGlyAsnIleIleMetPheAlaProThrLeuTrp  
AGACACACTCCAGTCAATTCTGGCTAGGCAACATAATCATGTTTGCCCCACACTGTGG  
TCTGTGTGAGGTCAGTTAAGGACCGATCCGTTGTATTAGTACAAACGGGGGTGTGACACC

8521 AlaArgMetIleLeuMetThrHisPhePheSerValLeuIleAlaArgAspGlnLeuGlu  
GCGAGGATGATACTGATGACCCATTTCTTTAGCGTCCTTATAGCCAGGGACCAGCTTGAA  
CGCTCCTACTATGACTACTGGGTAAAGAAATCGCAGGAATATCGGTCCCTGGTCAACTT

8581 GlnAlaLeuAspCysGluIleTyrGlyAlaCysTyrSerIleGluProLeuAspLeuPro  
CAGGCCCTCGATTGCGAGATCTACGGGGCTGCTACTCCATAGAACCACCTTGATCTACCT  
GTCCGGGAGCTAACGCTCTAGATGCCCCGGACGATGAGGTATCTTGGTGAAC TAGATGGA

8641 ProIleIleGlnArgLeuHisGlyLeuSerAlaPheSerLeuHisSerTyrSerProGly  
CCAATCATTCAAAGACTCCATGGCCTCAGCGCATTTTCACTCCACAGTTACTCTCCAGGT  
GGTTAGTAAGTTTCTGAGGTACCGGAGTCGCGTAAAAGTGAGGTGTCAATGAGAGGTCCA

8701 GluIleAsnArgValAlaAlaCysLeuArgLysLeuGlyValProProLeuArgAlaTrp  
GAAATTAATAGGGTGGCCGCATGCCTCAGAAAACCTGGGGTACCGCCCTTGCGAGCTTGG  
CTTTAATTATCCCACCGGCGTACGGAGTCTTTTGAACCCCATGGCGGGAACGCTCGAACC

8761 Gly  
ArgHisArgAlaArgSerValArgAlaArgLeuLeuAlaArgGlyGlyArgAlaAlaIle  
AGACACCGGGCCCGGAGCGTCCGCGCTAGGCTTCTGGCCAGAGGAGGCAGGGCTGCCATA  
TCTGTGGCCCCGGCCTCGCAGGCGCGATCCGAAGACCGGTCTCCTCCGTCCCGACGGTAT

8821 CysGlyLysTyrLeuPheAsnTrpAlaValArgThrLysLeuLys  
TGTGGCAAGTACCTCTTCAACTGGGCAGTAAGAACAAAGCTCAAAC  
ACACCGTTCATGGAGAAGTTGACCCGTCATTCTTGTTCGAGTTTG

FIG. 17-10

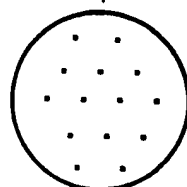
SUBSTITUTE SHEET

35 / 40

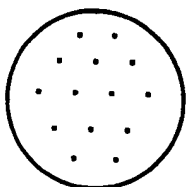
**IMMUNOLOGICAL SCREENING IN BACTERIA**

FIG. 18

Transform E coli with Recombinant Plasmids

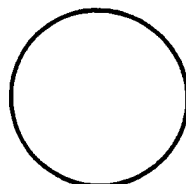
(Blot Bacteria on  
Nitrocellulose Filter)

IPTG Plate

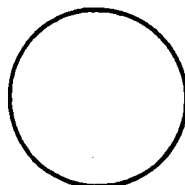


Lyse with Chloroform

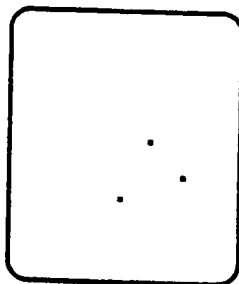
BSA absorbtion/DNAse/Lysozyme

Incubate with primary  
antibody

Wash

Incubate with  
<sup>125</sup>I secondary antibody

Wash



Autoradiograph

**SUBSTITUTE SHEET**

36 / 40

FIG. 19-1

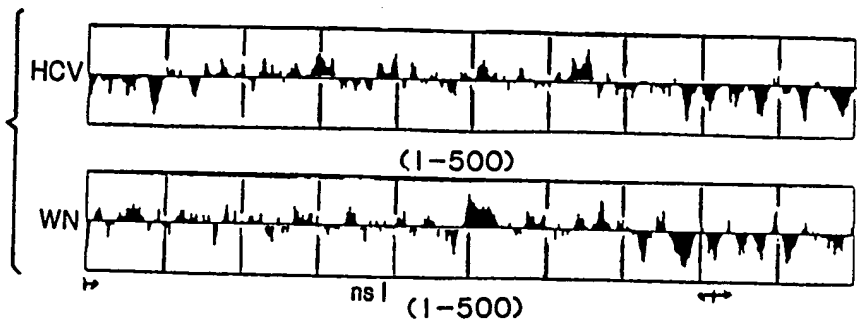


FIG. 19-2

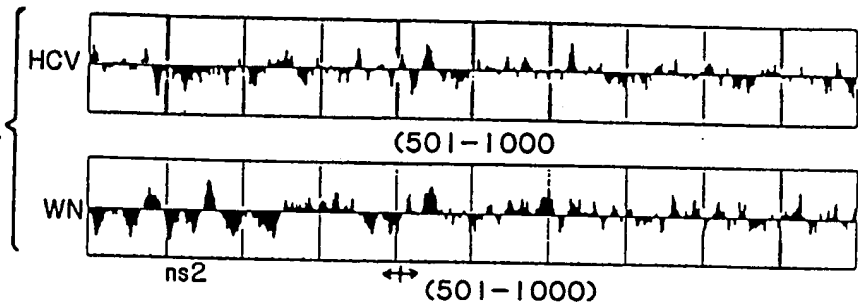


FIG. 19-3

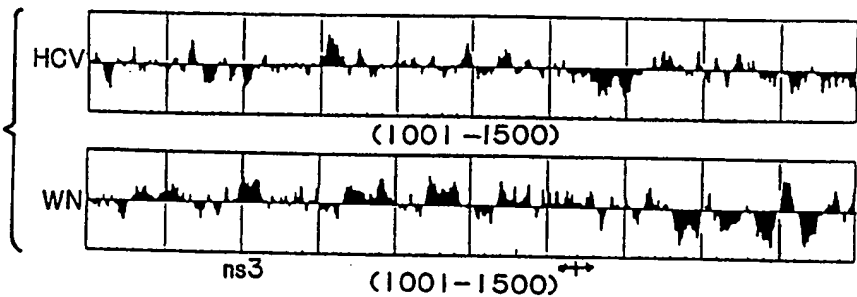


FIG. 19-4

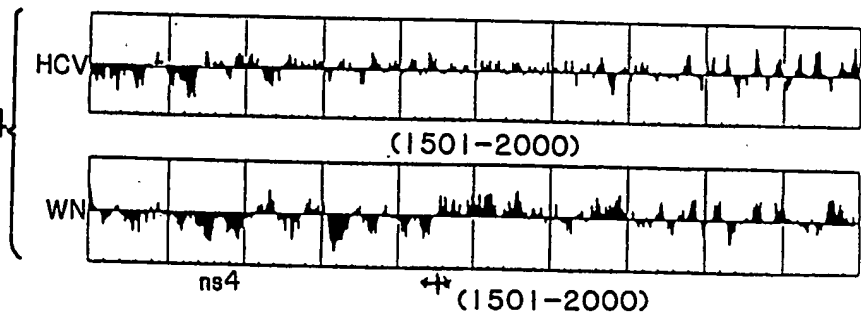
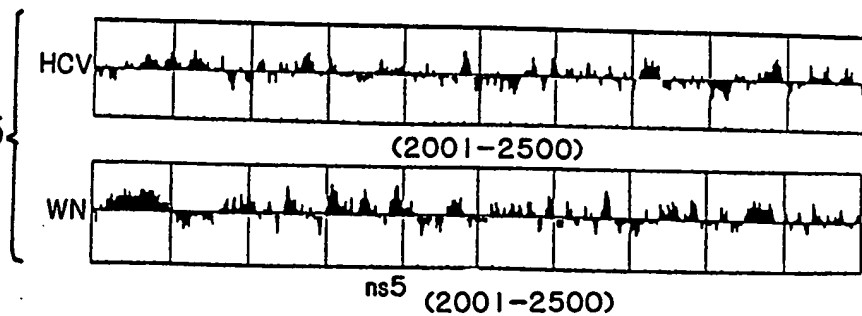
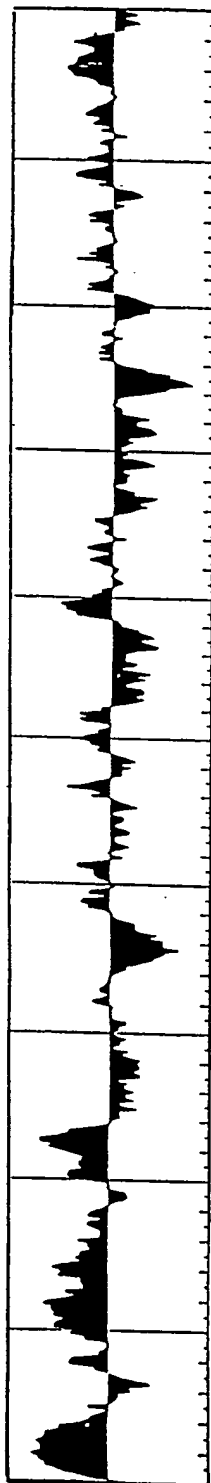


FIG. 19-5



SUBSTITUTE SHEET

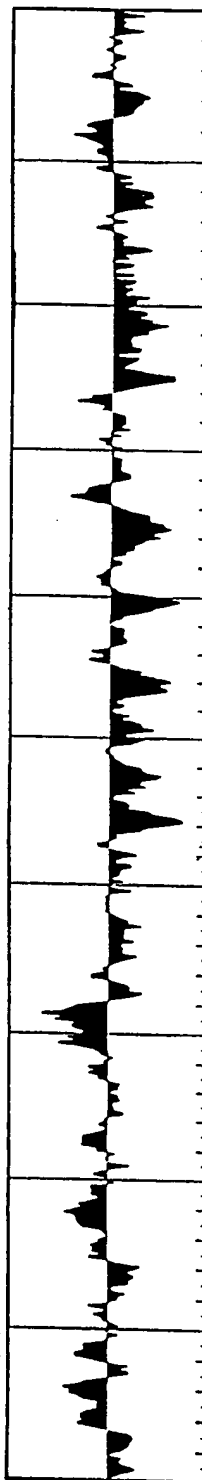
FIG. 20-1



ANTIG INDEX

HCV (1 - 500)

FIG. 20-2

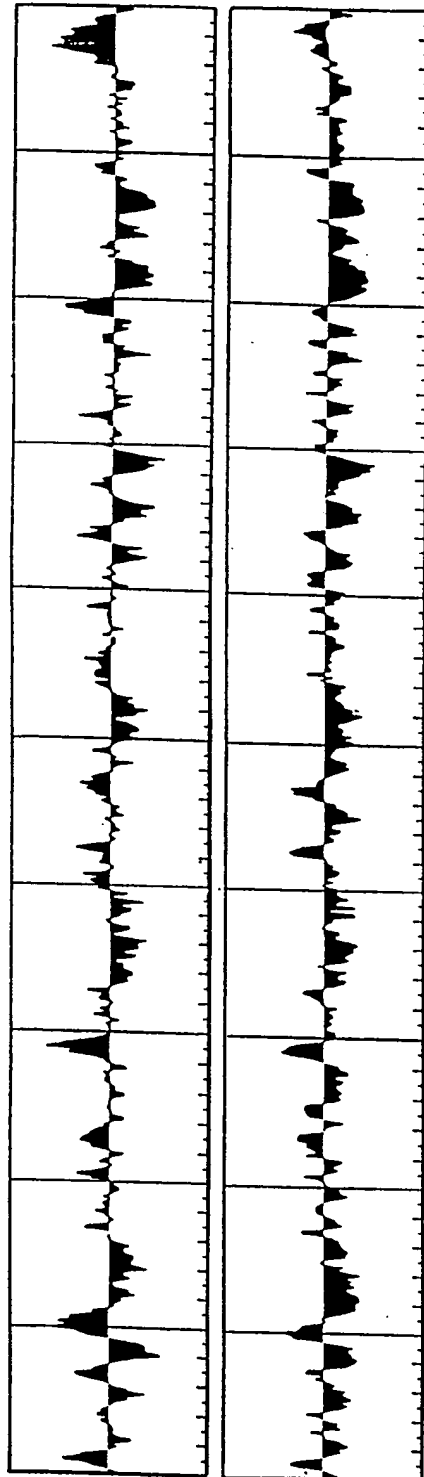


ANTIG INDEX

HCV (501 - 1000)



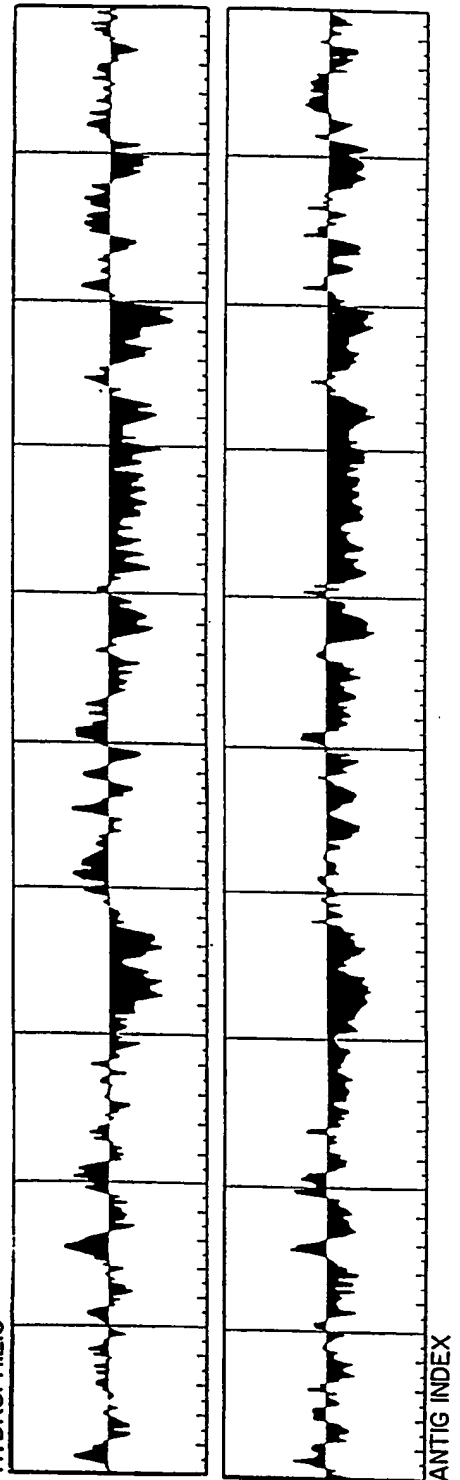
FIG. 20-3



ANTIG INDEX

HCV (1001 - 1500)

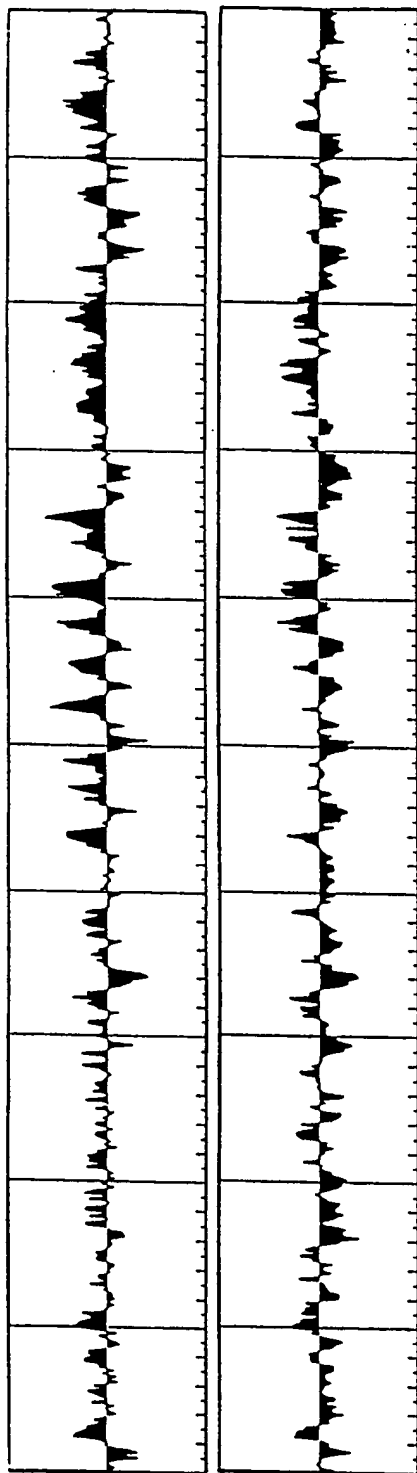
FIG. 20-4



ANTIG INDEX

HCV (1501 - 2000)

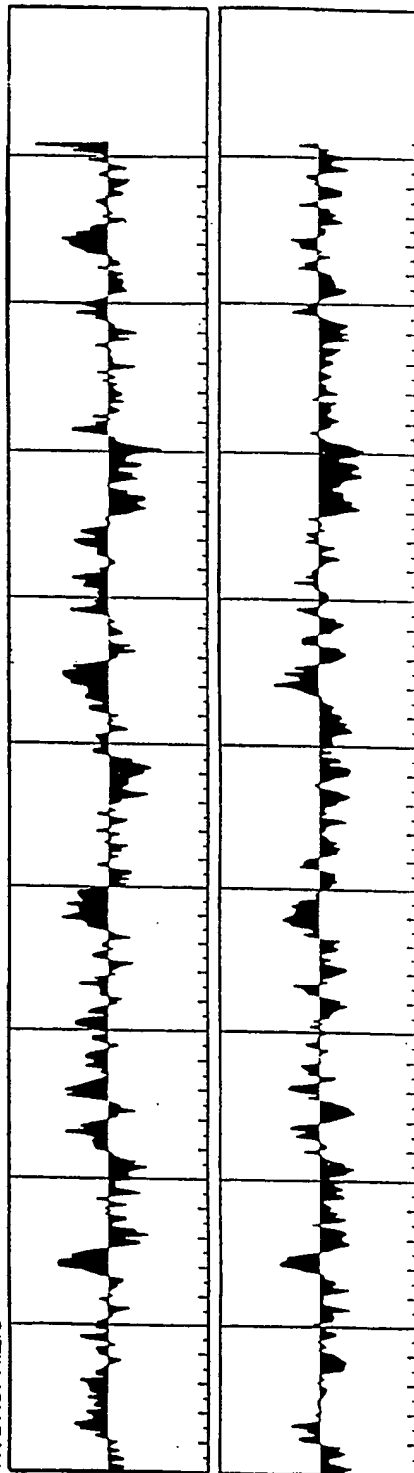
HYDROPHILIC FIG. 20-5



ANTIGEN INDEX

HCV (2001 - 2500)

HYDROPHILIC FIG. 20-6



ANTIGEN INDEX

HCV (2501 - 2955)

### Some conserved co-linear peptides in HCV & Flaviviruses

NS3 region

NS5  
Highly-conserved  
Polymerase  
region

Flaviviruses  
(Yellow Fever,  
West Nile, Dengue)

TATPPG-----SAAQRRGRIGRNP-----

-----GDDCVV

\*\*\*\*\* \* \*\*\*\*\* \*\*

\*\*\* \*\*

HCV

TATPPG-----SRTQRRGRTGRGK-----

-----GDDLTV

#1348 #1483 #2737

FIG. 21

# SUBSTITUTE SHEET

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/US90/01 348

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC IPC(5):A61K 39/12;C07H 21/04;C07K 7/06,08,10,13/00,15/28; US: 536/27;435/6,7,69.1,320,240.1;530/324-327,350,387,416;424/89		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
U.S.	536/27;435/6,7,69.1,320,240.1 530/324-327,350,387,416; 424/89	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched <sup>8</sup>		
Automated Patent Search, Chemical Abstract Service		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT <sup>9</sup></b>		
Category <sup>*</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
A,P	US, A, 4,870,026 WANDS ET AL. 26 September 1989	1-34
A	US, A, 4,702,909 VILLAREJOS ET AL. 27 October 1987	1-34
A	US, A, 4,542,016 TREPO 17 September 1985	1-34
X,P	Science, "Isolation of a c DNA clone derived from a blood-borne non-A, non-B viral hepatitis genome" volume 244, pp 359-362. CHOO ET AL. 21 April 1989 see abstract & figs.	1-12,15,16,18-20,22,28-31,33
X,P	Science, "An assay for circulating antibodies to a major ehologic virus of human nonA,nonB Hepatitis" volume 244, pp 362-364. KUO ET AL 21 April 1989 see abstract.	18,19,22
<p><sup>*</sup> Special categories of cited documents: <sup>10</sup></p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&amp;" document member of the same patent family</p>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
28 MAY 1990	13 JUL 1990	
International Searching Authority	Signature of Authorized Officer	
ISA/US	NINA COSSANNA	

## FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Continuation fo Classification:  
 IPC(5) C12N 1/11,15/02,51 C12P 21/02; C12Q 1/68;  
 G01N 33/53

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE <sup>1</sup>

This International search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers \_\_\_\_\_, because they relate to subject matter <sup>12</sup> not required to be searched by this Authority, namely:
  
2. ☐ Claim numbers \_\_\_\_\_, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out <sup>13</sup>, specifically:
  
3. ☐ Claim numbers \_\_\_\_\_, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING <sup>2</sup>

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.
2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:
3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:
4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

## Remark on Protest

- ☐ The additional search fees were accompanied by applicant's protest.  
☐ No protest accompanied the payment of additional search fees.